ERASMUS SCHOOL OF ECONOMICS

January, 2020

Prediction of market share of fast moving consumer goods via market share attraction model using parametric method.

A STUDY OF RETAIL-INDUSTRY PRODUCTS

Author: Aditya Sharma
Student number: 476867
Supervisor: Groenen, P.J.F.

Abstract: Fast moving consumer goods (FMCG) have always been a dynamic and innovative category acquainted with a variety of vital variables including consumers, promotions, logistics, supply-chain and market share. Companies like Unilever, Procter & Gamble, Henkel, Colgate-Palmolive, etc., are making use of advanced data-science and analytics, business intelligence and econometric modelling in analysing their products' portfolio and implementing more optimized business and marketing strategies in order to increase their profit margin, to expand their market share, and to overcome the challenges imposed by the competitors. In this paper, the researcher will predict the market share of various products under the fabric solutions category of the home-care segment. For the purpose of prediction, the market share attraction model (MSAM) will be implemented along with pooled and random effects model using the OLS method. In the end, the researcher will analyse the results and give his conclusion based on those results.

Keywords: FMCG, econometric methods, market share attraction model, competition.

Contents

1 Introduction

Consumer goods are those which are consumed by a consumer. These products are divided into three categories namely durable, non-durable and services. Most of the goods within the non-durable category are the ones that comes under fast moving consumer goods (FMCG). In this research study, the researcher will be working with the data related to FMCG products consisting of variables like sales, market share, volume and specific product-related benefits. Before proceeding to the main research question, the researcher would like to share some basic information, and provide answers to the following questions related with FMCG business. What is FMCG? How does a FMCG business works in the present era? What sorts of products are sold within FMCG? Who are the current leaders in the FMCG business?

FMCG consists of non-durable goods. Due to their fast moving nature, the products within FMCG are characterised by relatively low prices and a short shelf life due to which they are purchased and consumed rapidly by the public. Some common examples of FMCG products are detergents, ice-creams, soda, tea, cosmetics, toothpaste, drugs like paracetamol and much more. FMCG products usually forms a bulk part of a user's spendings on average. FMCG can further be divided into different divisions depending upon the different characteristic of the products that are being sold. The various divisions witihn the FMCG business are as follows:

- Processed foods: cereals, chips, cakes, biscuits, pies, pastries.
- Beverages: energy drinks, Cola, Pepsi, juices.
- Baked goods: cookies, croissants, and bagels.
- Fresh foods and dry goods: fruits,vegetables, raisins and nuts.
- Medicines: Aspirin, pain relievers, Paracetamol.
- Cleaning products: baking soda, kitchen cleaner and glass cleaner.
- Cosmetics and toiletries: hair care products, toothpaste, soap, shampoo, toilet cleaner.

FMCG is a really competitive business with a high turnover rate and a reliable source of revenue. Some of the big players in FMCG business are Unilever, Procter & Gamble, Nestle, Reckitt Benckiser, Coca-Cola, Johnson & Johnson and so on. Due to the highly competitive nature of FMCG business, companies look at and analyze every key area such as packaging, turnover rates, brand community, et., in order to increase their net profit and market share in various segments of products across different markets. Packaging plays a vital role in the production process. Primary packaging handles the responsibility of product protection and shelf life while simultaneously providing information and sales incentives to consumers.

In the past 6-8 years, the e-commerce business has grown exponentially and FMCG companies have used this platform to their advantage in order to increase the sales of their products, provide lucrative deals and discounts to consumers and thus increase the revenue and market share of their business. The consumer procure a number of benefits from making online purchases such as convenient delivery of their products right at their door-steps, a broad range of categories of products to select from, as well as low−priced goods whose prices are even lower than that of local grocery stores and supermarkets. The online platform of selling non-durable products is growing day by day as companies seek to increase the efficiency of delivery logistics which shortens the delivery time. There is a one−to−one correlation between the buyer who is searching for the product online and purchase since most of the times buyers are certain about the product they want to buy. Also, non-durable products have lower online browse−to−buy intention than durable ones, but they still present a strong browse−to−buy correlation, which may be a factor in their increasing online sales. As mentioned in Victoria and Ganesan (2014) e-commerce platform is a rich source of information about the consumer and their likely response to online sales promotion activities. Accessing the information on how the consumers reacts and behaves to different price and display promotions from time to time would be really helpful for planning future sales promotion activities. In order for the FMCG companies to build trust and commitment, their online sales should tap preferences and perceptions of consumers.

In this modern era of vigorous competitiveness, it is crucial for a company to keep up with the pace of their respective markets. In order to understand the ever-evolving market and the customers within it and to keep their customers intact and happy, it is vital for companies to identify and understand different variables that affects their market share and act accordingly. Thus making sure to adapt and evolve with the changing markets' outlook. Kotler and Armstrong (2010) mention that for marketing, retaining and growing current customers is as important as acquiring new ones because the former method leads to sustainable customer lifetime value and the latter one brings new sources of values to the company. These two processes of 'retaining' and 'growing' of customers plays a huge role in the growth/decline of market share of a given brand of a product. Furthermore, in this era of big data, each company has access to an abundance of data and thus by analysing the various important variables like promotional sales, market share, products-specific benefits, pack size, demographics, rate of sale, volume on discount and so on, these companies can plan the next move for their market strategies.

This paper will focus on the above mentioned variables that represent the performance of products within a given market. Using these variables the researcher will be estimating as well as predicting the market share of fabric solution products via different econometric methods.

The variables present in the data-set are known as the key-performing indicators (KPI's) and using these KPI's along with the market share attraction model (Nakanishi and Cooper, 1988; Fok and Franses, 2004), the researcher will perform the desired analysis and prediction of the market share with respect to a benchmark competitor. One of the important questions in this research study is, why do we need market share attraction model? Why is it so much beneficial while dealing with multi-variate data-sets?

First, brands within a given product segment are obviously related. Second, is the concept of competition which is very well explained and implemented using the market share attraction model. Here competition also includes the role of promotion, display effect in the market which basically leads to increase in a brand sales by drawing sales from competing brands. Third, market share attraction model helps us to study and understand the market structure, how does brands compete with each other and also helps in measuring cross elasticities across different variables. Even if there is no presence of cross effects, considering and taking into account all the equations under the market share attraction model is much more beneficial in terms of high efficiency. One of the important advantages of using the concept of market share as opposed to sales is that market shares is relative market performance measure and can be compared across different geographical regions whereas sales are sometimes not stationary as in the case of category expansion or seasonal movements. That is why sales model is comparatively weaker and inconsistent when it comes to doing predictions regarding how well a brand is performing in the market. For this particular research study which includes FMCG brands, market share attraction model acts as a catalyst when it comes to evaluating the effects of different marketing variables like promotional sales, volume sales, etc., on the overall market share of a given brand of products within fabric solutions' category being conditioned on the competitors' responses.

In terms of other econometric models, the researcher will start by performing the simplest linear regression (OLS) in panel data, called the pooled ordinary least squares (OLS). Then, the researcher will proceed by performing the 'random-effects' regression model on panel data in order to analyze how well the model with respect to the given FMCG fabric-solution products' data performs as compared to other models.

Also, there might be some endogeneity present in the data. For example, the trust of consumers in a given brand of a product can be considered as an individual effect and it might be correlated with the volume sales per period of same product. For example, in the Netherlands Robijn has been the biggest and most trustworthy local detergent product for many years and this individual effect of trustworthiness might be included in the error term, which in turn might be correlated with the volume sales of Robijn brand (Setyawan and Imronudin, 2015; Fianto et al., 2014). Also, there is a belief that one of the variables 'market share' of a brand is highly persistent and is correlated with its own lagged values as well as the market share of other brands. For example, a discount promotion implemented in a given time-period can still have its after-effects in the sales growth of a given product in successive time periods and simultaneously affect the sales growth of other brands in the present as well as future time-periods. (Pauwels et al., 2002; Kopalle et al., 1999).

The main aim of this paper is to predict the market shares of FMCG fabric solution products with respect to a benchmark product. In order to evaluate the market shares with respect to a benchmark product, the market share attraction model will be implemented using parametric method. Prediction is done by executing various econometric methods such as OLS, randon-effects, time-series, simulations, etc., in order to obtain the estimated values of the relative market shares. In this research study, the market share attraction model (MSAM) along with pooled and random effects model are implemented by the researcher to obtain the final results. In the end, the researcher will compare the results obtained from these models to investigate which econometric model gives the most accurate and precise results.

The remainder of this paper is organized as follows. First, the researcher will discuss in detail the previous conducted research studies, relevant papers from different journals and articles in presenting the role and functionalities of different methods, their advantages/disadvantages with respect to the current data-set under the Section 2. Under the Section 3, the researcher will describe the products that has been used in the data-set depending upon the demographics, pack-size, category, segment, product-specific benefits, promo sales, volume sales and the market share. Under the Section 4, the researcher will discuss in detail different econometric methods such as OLS, random effects and MSAM, their implementation, and how these methods have been useful in estimating and predicting the dependent variable. The key findings from this study will be presented under the Section 5. Under the Section 6, the researcher will conclude the main findings from this study in a sterling-concise manner and lastly, under the Section 8, a further discussion on the future scope of research in the field of analyzing the data related with FMCG products and non-parametric methods will be discussed.

2 Literature review

One of the initial instances of econometric modelling being introduced and utilised for data analysing purpose was carried out by Jan Tinbergen in Tinbergen (1940). It laid the foundation for many interesting and renowned research studies to be performed in the coming years in various fields such as machine learning, data-science and artificial intelligence.

Data can be available in many different forms. One of the most common forms of data structures is panel data which contains observations across different individuals over a given period of time. It combines both cross-sectional and time-series data. Hsiao (2007) presents the advantages/disadvantages that a person may face with while working with panel data. Blossfeld et al. (2009) uses panel data in order to obtain useful information about learning environments and to apprehend the impact on educational choices because of parallel processes.

In this research study, the researcher is going to make use of MSAM for the given data-set. Market share attraction model is effective while working with panel data as it integrates the dynamics specification within its working criteria with great precision and accuracy thus producing efficient estimates and reliable results. One of the earliest theoretical implications of this model can be found in Kotler (1965) where the model was used for simulation study of marketing strategies for competitive duopoly. Nakanishi and Cooper (1974) introduced a theoretical concept for accurately estimating the parameters in market share model. In order to attain asymptotically efficient estimates of the parameters in the market share attraction model, Bultez (1978) make use of various approaches to linearize the market share attraction model. Simon (1978) makes use of the simulation study of Kotler (1965) in order to carry out its own in-depth analytical research and provides interesting and valuable results, such as, the model produced changes in price elasticity over time, which do not appear to be economically plausible. In order to better understand the predictive power of MSAM as compared to the other classical market share model, a detailed research study was carried out by Naert and Weverbergh (1981) and this particular case study entrusted me with an inspiration for my own research topic using a unique data-set of fabric-solution products.

Several available methods and techniques utilize panel data to perform constructive and efficacious analysis regardless of the data-set. Among the various methods already available, pooled OLS and random-effects model are the ones that make use of the panel data as a desired form of data-set. In the pooled OLS method, the individual-effect is taken as constant across individuals and different time-periods. Similarly the effect of the explanatory variable across individuals and time-periods is constant as well. The advantages of using pooled OLS is that this method is highly parsimonious and computationally very simple. The only disadvantage of this method is that no heterogeneity is allowed in the model. In the random-effects method, we take into account that there is a presence of individual effect in the model and this individual effect varies across individuals. The individual effect is not correlated with the explanatory variables under random-effects specification. The advantages of using random-effects method is that it is computationally simple and it allows heterogeneity to be present in a given model. Random-effects estimators also allows the estimation of time-invariant regressors.

In recent times, a lot of research studies have combined the use of panel data with either pooled OLS or random-effects. The paper Balestra and Nerlove (1966) revealed the benefits of utilising the pooled model regression for a dynamic data-set which leads to more consistent and accurate estimates as compared to classical OLS. A research study on riverbed permeability and groundwater conductivity was carried out using the pooled data in Calver (2001) in which the researcher was able to acquire some useful results and conclusions related with hydrological and hydrogecological assessment where specific investigation was not possible. One of the interesting topics in which a researcher compared the political economy of various countries comprising of different key performing indicators (KPI's) in order to derive valid inferences from statistical comparisons was achieved via the help pooled data modelling Kittel (1999). When it comes to presence of dynamics specification in data, nothing gives more accurate estimates than pooled modelling techniques which has been further explained in-depth by Baltagi (2008).

Random effects model allows the individual effects to be included in the model but these individual effects are not correlated with the explanatory variables which will be further explained in depth in the Section 4.3.2. In the paper Hogeweg and Hesper (1990), the researcher showed the benefits of the inclusion of the individual-effect within a model as the model was able to extract a real-sensible meaning from the given data-set in this particular research study. In the paper Dingemanse and Dochtermann (2013) the researcher quantified the individual effects in order to get a standardized measure of individual variations and included these variations in his model and explained how these variations affected the overall results in the end. The paper Greene (2005) examined on how measuring the effects of heterogeneity can lead to better and accurate estimates. In the study Berkey et al. (1995), random effects modelling was found to produce less biased results and performed really well because of the inclusion of covariates that explained the heterogeneity. Justifying appropriate consistency and optimality with respect to a given data-set was achieved via random effects model in Ma et al. (2003). This was implemented using an orthodox best linear unbiased predictor approach. There can be some inconsistency while working with random-effects approach and it can produce misleading results for evaluation purposes which is explained thoroughly in Fieuws and Verbeke (2004).

3 Data

The raw data for the research purpose is available from Nielsen database from where the fabric-solution products' data of four different European countries; Germany (DE), France (FR), Spain (ES) and United Kingdom (UK) is selected. In this research study, the above mentioned four countries combine to form a large market within EU and the researcher will define this market as 'M'. One of the reasons to choose these four countries is that they are the biggest home-care products consuming countries, which in turn, forms a big chunk of revenue source for big FMCG companies. The various key performance indicators (KPI) included in the raw data-set are: value sales, volume sales and promo sales. The other explanatory variables present in the data are: company description, global brand, local brand, segment, format, variant, actual pack size and product-related benefits. There are a total of 705 products listed in the raw data-set which are spread across 38 time-periods starting from P8 2016 till the latest updated period P6 2019 in terms of promo sales, value sales, volume sales and market share. In Nielsen database, each period is defined as a cluster of four weeks starting from the month January of each year till December end, meaning that, in a given calendar year there are 13 periods. We have a data-set of 38 periods, equalling to three years worth of data. There is also product-specific benefits data that is included in our raw data-set which comprises of benefits such as care, eco, fragrance, sensitive, whiteness, baby, stain removal and so on. These benefits variables are time independent categorical variables.

3.1 Data preparation

There can be inconsistency and/or presence of errors while working with real world data. Data preparation is an initial step which lays the foundation of a concrete final base (format) in which the researcher can transform and utilise raw data into an understandable format. Data preparation is a proven method of resolving such issues and arrange raw data for further processing.

The raw data set was downloaded from Nielsen database which comprised of 3000 products from 12 different countries but from those 3000 products and 12 countries, the researcher only choose to work with 705 products and 4 countries and the reason of choosing such specification is already mentioned in Section 3. Also, the market share variable was absent from the original data set which the researcher calculated by dividing the sales of each product within a given time period by the total sales of market 'M' in the same time period. Since the whole process of using market share attraction model depends upon using a benchmark product for calculating the relative market shares of all the products, a separate column containing the market share value of the chosen benchmark product was created in the main data-set. The chosen benchmark product will be explained further in detail in the upcoming paragraphs of Section 3.

One of key areas to be very careful about while performing data preparation is the handling of categorical variables. Categorical variables are those which are discrete and are not continuous. In this particular research study, the benefits related with each product are the categorical variables. In the raw data set the benefits related data was just provided as separate column in which the benefits of each product was mentioned by their name in front of the respective products. The researcher created separate columns for each benefit in the main data set and they contained the values one if a given product contains that benefit and the value zero if the given product does not contain that benefit.

Also, the MSAM is implemented by using the log transformation. So, the researcher converts all the values in the market share and the benchmark variables into logarithm values in the main data set. By implementing all the points mentioned above, the main data set is finally completed and is ready to be used for the modelling and analytical purposes.

3.2 Descriptive stats

The Table 1 presents the percentages of the products that falls under different KPI variables and provide useful information regarding variables such as market description, company, segment, format and so on. As you can see from the Table 1e, approximately 23% of products have 'care' benefit associated with them whereas 77% does not have this benefit at all. Similarly 29% of products have 'stain removal' benefit associated with them. The stats associated with the mentioned benefits exhibit the purchasing behaviour of the customers when they see a benefit associated with a product as compared to a product with no benefits (mentioned on the pack) at all. The magnitude of the purchasing behaviour changes drastically when more than one benefit is mentioned on a product but this point falls outside the scope of the current research and will not be discussed further in this study. These benefits represent whether the product is specified for a care purpose in case of 'care' benefit or whether it is specific for the protection of colouring of clothes under the 'colour' benefit or whether the product is specific to make the clothes smell fresh with nice fragrance under 'frag fresh' and so on.

As you can see under the Table 1a, UK has the least number of products under its portfolio whereas Spain has the maximum with a market coverage of 45.5% . One of the reasons for UK to have a lesser diverse portfolio is because of its size as compared to the other three countries in addition to the strong market presence enjoyed by Unilever in fabric-solution segment in the UK. This makes penetration by competitors further difficult into a market already characterised by strong base of loyal customers of Unilever along with the presence of less costlier local-detergents which are also popular among the average-income class of people. Whereas in Spain, there is no clear favourites among the FMCG companies to be in a strong-hold position although Procter & Gamble has a strong-hold position in the Spanish market, the market is receptive towards other brands and lets the consumers decide the fate of the product. Usually the price, discounts and benefits related with the products plays a huge role in motivating people to try out different products even if they are new in the market. Germany also has a low diversity in its portfolio because of the stronghold of Henkel in German market.

Under the Table 1d, you can see various companies playing in the FMCG fabric segment in the four chosen countries. 'Others' consists of small local brands which are area-specific depending upon the needs of local people. These brands as individuals, usually have a very small portfolio and cover a small share in the market which is why they've been clubbed under 'Others'. The company having the largest share of products in its portfolio is Henkel owing to its stronghold in Germany. Next is Procter & Gamble leading its way in FR and ES followed by Unilever with 89 products under its portfolio due to its stronghold in UK market.

The detergents nowadays is sold under different formats namely capsules, liquid cons., paste, powder and so on. Paste has been phased out from the market meaning most of the big FMCG have stopped the production of paste detergents which can be easily seen from Table 1c. Due to the environmental issues and the need to make detergents more environment friendly, these companies are trying to move away from powder format as well. Inspite of rigorously promoting the other formats, powder still remains the second largest division 30% under which the detergents are currently sold as seen from the Table 1c. Semi-cons. (basic liquid detergents) are considered to be less harmful than powder detergents and currently are the largest sold format of the fabric-solution segment. Liquid cons., an upgraded and concentrated version of semi-cons. performs at par with the latter. The internal gross margin of liquid cons. is higher than that of semi-cons. They occupy a small share in the current market but are expected to be the leader of fabric-solution category in the future. Capsules which has a market share of 11.2% is also one of the formats which is perceived well by the consumers as it is easy to handle and use. Procter & Gamble was the first one to introduce this format on a big scale in the market and the first company to patent this design on a large scale in this category.

As seen from Table 1b, segment variable represents that 92% of the products are made for the regular washing task whereas the remaining 8% is made for the delicate washing of woolen fabrics or other fabric clothes which needs extra care during washing.

Global brands includes the products which are available on a global scale and represents the main name under which the FMCG promotes and advertises the 'brands' division which in-turn, is a subcategory of Global brands. Private labels which is a combination of products developed by discounters group like Lidl, Albert Heijn and so on, covers 15% of the market share. Ariel which is the global brand of Procter & Gamble is next on the list with 10.5% coverage of the market. For Unilever, the largest global brand product is 'Dirt is Good' which covers 8.5% of the entire market. The remaining global brands covers very small portions which falls in the range of 0.5% to 4% .

Brands division consists of products that are sold in country-specific regions or those products which are sold under different names in different regions or country. As you see from the Table 1g that Ariel is also promoted as a brand under the main division of 'Global brand' and thus it covers exactly the same market share of 10.5% . Private Labels again tops the list with

15% share and the next brand (excluding ALL OTHERS) is Skip with 7.4% share.

As mentioned in the Section 1, the dependent variable is the market share of various products with respect to a chosen benchmark product across the market 'M'. The chosen benchmark product for this research study is from Henkel company with global brand name as 'Wipp Express' and the format of the product is semi-concentrated liquids. The actual pack-size of this product is 2914 ml. The main reason of choosing this product is that this is one of the products which has been available for almost all the 38 periods and is performing really well in terms of sales for the last one and a half year. In this competitive market where new products are launched every week, where products come and go from the market like they never existed, for a product to maintain a good reputation and sales for most part of 38 periods suggests the stability and trustworthiness of product within the market and gives the researcher enough motivation to perform the respective modelling and predictions with respect to this product. The explanatory variables utilised in this research paper are promo sales, volume sales and market share of fabric solution products along with the benefits related categorical variables such as care, frag fresh, stain removal, colour and so on.

From Table A2 we can see the mean of the promotional sales were around Euro 40,000 to 50,000 at the end of 2016 which increased at a booming rate to its peak point of around Euro 80,000 to 1,00,000 in 2017 pointing to the fact that a lot of new fabric products were launched during this period and these products needed a push in the market in order to catch the attention of the consumer. So, a lot of these newly launched products were sold on promotions and in order to compete with this increase in promotional sales, the already established products were also forced to increase their sales under promotions in order to maintain their market share. But in the late periods of 2018 till present period of 2019 we saw a decline in promotional sales to a point of around Euro 30,000 to 40,000 which shows that most of the newly launched products of 2017 have established themselves in the market and thus their dependency on promotions in order to attract the customers have decreased.

In the above paragraph, the researcher has mentioned that the year 2017 represents the launch of various new fabric solution products leading to high promotions intensity, which is further backed by the stats of increased volume sales (mean values) in the the year 2017 as shown in Table A3. As seen in Table A3 the decrease in the volume sales from midway year 2018 till P07 2019 supports two points. First, there is a decrease in the intensity of the promotions. Second, there is a stability in the sales volume of products which can be due to decrease in promotions intensity along with different factors; example people becoming aware of the harm caused due to the over-utilisation of detergents on environment.

In Table A1 the market share of the of best-performing products in a given period is always between 35% to 45% apart from few exceptional periods like 17P12, 17P13, 18P04 and 18P12 where the market share is above 65% and periods like 17P05, 17P08, 18P05 and 19P01 where the market share is below 18% for the best performing products.

Table 1: **Key performing indicators description**

(a) Market description

Countries	Tot	%
Germany(DE)	113	16.02
Spain (ES)	321	45.53
France (FR)	159	22.55
United kingdom (UK)	112	15.88

(b) Segment

Division	Tot	%
Delicate Wash	56	7.94
Main Wash	649	92.05

(c) Format

Division	Tot	%
Capsules	79	11.20
Liquid Concentrated	61	8.65
Paste	1	0.14
Powder	209	29.64
Semi-Cons/Std Liquids	330	46.8
Tablets	25	3.54

(d) Company

Name	Tot	%
C-P	1	0.14
Henkel	154	21.84
Others	209	29.64
P&G	101	14.32
Private Labels	106	15.03
R&B	45	6.38
Unilever	89	12.62

(e) Benefits

Benefit	Yes	No	% (Yes)
Care	160	545	22.7
Eco	27	678	3.8
Frag fresh	101	604	14.3
Frag perfume	40	665	5.6
Hygiene	2	703	0.2
Sensitive	47	658	6.6
Stain removal	205	500	29.1
Whiteness	27	678	3.8
Color	79	626	11.2
Baby	5	700	0.7

(f) Global brands

Main prod.	Tot	%
Ariel	74	10.49
Dash\Bold	15	2.12
Daz	5	0.70
Dirt Is Good	60	8.51
Dixan	21	2.97
Fairy	5	0.70
Le Chat	10	1.41
Lenor\Downy	1	0.14
Mir	16	2.26
Others	266	37.73
Persil (Henkel)	20	2.83
Persil (Unilever)	7	0.99
Perwoll	5	0.70
Private Labels	106	15.03
Radiant	2	0.28
Spee	5	0.70
Super Croix	8	1.13
Surf	18	2.55
TP	2	0.28
Weisser Riese	16	2.26
Wipp Express	26	3.68
Woolite	7	0.99
Xtra (Henkel)	10	1.41

(g) Brands

Prod.	Tot	%	Continued...		
Actiff	1	0.14	Micolor	14	1.98
All other	61	8.65	Mir	16	2.26
Ariel	74	10.49	Norit	13	1.84
Bionicdry	1	0.14	Om bianco	6	0.85
Bold	11	1.56	Omo	13	1.84
Burtt	2	0.28	Perlan	3	0.42
Colon	26	3.68	Persavon	1	0.14
Coral	2	0.28	Persil	20	2.83
Dalli	7	0.99	Persil (UNI)	15	2.12
Dash	4	0.56	Perwoll	5	0.70
Daz	5	0.70	Pri label	106	15.03
Detersolin	9	1.27	Punto matic	10	1.41
Dixen	21	2.97	Rainett	3	0.42
Domal	1	0.14	Rei	2	0.28
Ecover	7	0.99	Rube	2	0.28
Eigenmarke	21	2.97	Sanytol	2	0.28
Elena	11	1.56	Sentimat	1	0.14
Fairy	5	0.70	Simply	1	0.14
Flota	7	0.99	Skip	52	7.37
Frosch	10	1.41	Soupline	1	0.14
Genie	3	0.42	Spee	5	0.70
Goir	3	0.42	Sunil	7	0.99
Halo	1	0.14	Super croix	8	1.13
Larbre vert	8	1.13	Surcare	1	0.14
Le chat	10	1.41	Surf	7	0.99
Lenor	1	0.14	Tandil	5	0.70
Luzil	13	1.84	Vizir	1	0.14
M verte	1	0.14	W riese	16	2.26
xtra	10	1.41	Woolite	7	0.99
Wipp express	26	3.68			

Note: The Table 1 presents the information regarding the variables that are present within the data-set of fabric solution products. Market description variable presents the info regarding the number of products available in a given country. Segments shows how many products falls under the delicate and main wash division. Format shows the info about the products available in different forms. Company variable is a indicator of number of products owned by a respective company within the market 'M'. Benefits variable tells us which are the most frequent and common benefits being associated with different products. It also presents the information about those benefits that are being under utilised in the present market. Global brands and brands gives us the information regarding the names of products under which they are advertised and promoted globally and locally respectively. It also showcase the composition of those brands' products within the present data-set.

[1] Tot:represents the number of products related with respective variables in a given table.
Total number of products: 705

4 Methodology

In this section the researcher will look in detail the methods that has been used for this research study and to understand the advantages of using the market share attraction model in predicting the relative market shares of fabric solution products. In the end, the researcher will summarise which of the given methods, pooled OLS, random effects or MSAM gives us more accurate predictions.

4.1 Market share attraction model

Market share attraction models estimate the own-effects and cross-effects of marketing-mix variables conditional on the competitive response (Karnani, 1985; Franses and Montgomery, 2002). As such these estimates can be also used to measure the market share performance given the competition. Nakanishi and Cooper (1988), not only discussed the theoretical properties of this model but also successfully showcased the implementation of this model using marketing data-sets in dynamic and non-dynamic specifications.

Market share attraction model takes into account the following points (Nakanishi and Cooper, 1988; Franses and Montgomery, 2002):

- Competition: What the competitors are doing.
- Description: Explaining why and what are the things happening in the market.
- Prediction: Providing info on how the market is going to behave in future.
- Profit oriented: Shows the practicality of using this model in real life based cases.
- Logically consistent: which satisfies:

$$0 \leqslant M_{it} \leqslant 1 \qquad \textstyle\sum_{i=1}^{\infty} M_{it} = 1, \tag{4.1}$$

where M_{it} refers to the market share of individual brand i at time period t.

One of the important concepts within the market share attraction model is **Attraction**. It is defined as the determinant of market share that the consumer perceives for each and every brand present in the market. The concept of attraction is the core of the entire market share attraction model. It was first introduced in Bell et al. (1975) and consists of the following **axioms**:

- $A_{it} \geqslant 0, \quad \sum_{i=1}^{\infty} A_{it} > 0,$
  where A_{it} refers to the attraction of individual brand i at time period t.
- $A_{it} = 0 \Rightarrow M_{it} = 0$
- $A_{it} = A_{jt} \Rightarrow M_{it} = M_{jt}$
- If $j \neq i$ then a change in M_{it} due to change in A_{jt} does not depend upon j.

So, from the above axioms we conclude:

$$M_{it} = \frac{A_{it}}{\sum_{j=1}^{\infty} A_{jt}} \qquad i = 1, ..., I. \tag{4.2}$$

4.1.1 Basic attraction model

Basic attraction model also known as multiplicative competitive interaction model is a simple version of fully extended market share attraction model in which the attraction of brand i at time t is given by (Naert and Weverbergh, 1981; Nakanishi and Cooper, 1988; Hanssens et al., 2003):

$$A_{it} = \exp(\mu_i + \epsilon_{it}) \prod_{k=1}^{K} x_{kit}^{\beta_k}, \tag{4.3}$$

where

μ_i : constant attraction for a given brand i,

x_{kit} : marketing instrument k for brand i,

β_k : the effect of marketing instrument k,

ϵ_{it} : attraction which is unexplained.

Some of the important assumptions within this model are (Nakanishi and Cooper, 1988; Franses and Montgomery, 2002; Hanssens et al., 2003)

- effect of marketing instruments on attraction is same for all brands,
- cross effects of instruments on attraction is not present.

Some of the advantages of using this attraction model are (Nakanishi and Cooper, 1988; Franses and Montgomery, 2002; Hanssens et al., 2003):

- As can be seen from 4.1, the market share of each brand is always between zero and one and the total market share sums upto one.
- The concept of partial effects regarding the changes in marketing instruments works reasonably well within this model.
- Price elasticity is easy to calculate and work with within this model.

One of the drawbacks of this model is that is does not follow the practical and real life based scenarios of day to day markets as it does not allow cross effects and specific brand effects to be present within the model. These cross effects and individual brand specific effects are present in the fully extended attraction model, also known as the market share attraction model in which the lagged marketing instruments and lagged market shares are included.

4.1.2 Market share attraction model (without dynamics)

The market share attraction model which includes the cross effects and brand specific effects is given by (Monahan and Sobel, 1994; Franses and Montgomery, 2002):

$$A_{it} = \exp(\mu_i + \epsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}} , \quad (4.4)$$

where the attraction of brand i is not only dependent upon its own marketing instruments but is also dependent upon the marketing instruments of the remaining $I-1$ brands. β_{kji} describes the effect of the marketing instrument k of brand j on the attraction of brand i . The market share attraction model is troublesome and laborious to work with in its above mentioned form so in order to make it easier to work with we have to linearize the above equation. This step of linearization of the model has some other advantages as well. These advantages are (Franses and Montgomery, 2002):

- Information regarding the behavior of an individual point w.r.t its neighboring equilibrium points is well presented and documented by the step of linearization.
- Non-linear equations can be easily approximated and their behavior can be governed by the set of linear equations. These linear equations are easy to work with and easy to solve in comparison to non-linear equations.
- Investigation about the stability of a point is done smoothly via the linearization process.

In order to understand and calculate how the market share of a given brand is performing within a given market we need to take a benchmark product from the given market and then with respect to that given benchmark product we can calculate the relative market share of that brand which is shown as follows. Taking this benchmark product also helps to solve the 'identification issue' problem as we know that if $I-1$ shares are known then the I-th share is known as well. The equation is written as follows:

$$\frac{A_{it}}{A_{It}} = \frac{\exp(\mu_i + \epsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}}}{\exp(\mu_I + \epsilon_{It}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kjI}}} \quad (4.5)$$

where

$$\frac{M_{it}}{M_{It}} = \frac{A_{it}}{A_{It}}$$

Taking logs,

$$\ln M_{it} - \ln M_{It} = (\mu_i - \mu_I) + \sum_{j=1}^{I} \sum_{k=1}^{K} (\beta_{kji} - \beta_{kjI}) \ \ln x_{kjt} + \epsilon_{it} - \epsilon_{It} , \quad (4.6)$$

where

$$\mu_i - \mu_I = \mu_i^*, \qquad i = 1, ..., I-1$$

$$
\begin{aligned}
\beta_{kji} - \beta_{kjI} &= \beta^*_{kji}, & i = 1, ..., I-1 \\
\epsilon_{it} - \epsilon_{It} &= \eta_{it}, & (\eta_{1t}, ..., \eta_{I-1,t})' \sim N(0, \Sigma^*)
\end{aligned}
$$

where Σ^* is a $(I-1) \times (I-1)$ covariance matrix.

The market share attraction model contains a lot of parameters. We might try out imposing some restrictions within this fully extended attraction model depending upon the kind of data present with us, and also what kind of market and competition are we expecting within a given geographical area. There are 3 different types of restrictions that can be imposed (Franses and Montgomery, 2002):

i) Restricted competition: in this scenario the attraction of a given brand only depends upon its own marketing instruments and explanatory variables.

ii) Restricted effects: this scenario is same as restricted competition with the added condition of having equal parameters across different brands.

iii) Restricted covariance matrix: defined as a covariance matrix in which there is no correlation between the attraction shocks.

The specification for the restricted models can be written as follows (Franses and Montgomery, 2002):

a) Restricted competition: The respective attraction model and reduced form of market share attraction model are written as follows:

$$
A_{it} = \exp(\mu_i + \epsilon_{it}) \prod_{k=1}^{K} x_{kit}^{\beta_{ki}},
$$

$$
\ln M_{it} - \ln M_{It} = (\mu_i - \mu_I) + \sum_{k=1}^{K} \beta_{ki} \ln x_{kit} - \sum_{k=1}^{K} \beta_{kI} \ln x_{kIt} + \epsilon_{it} - \epsilon_{It}.
$$

b) Restricted effects: The respective attraction model and reduced form of market share attraction model are written as follows:

$$
A_{it} = \exp(\mu_i + \epsilon_{it}) \prod_{k=1}^{K} x_{kit}^{\beta_k},
$$

$$
\ln M_{it} - \ln M_{It} = (\mu_i - \mu_I) + \sum_{k=1}^{K} \beta_k (\ln x_{kit} - \ln x_{kIt}) + \epsilon_{it} - \epsilon_{It}.
$$

The three types of restricted models, and their number of parameters and restrictions is for a convenience summarized in the Table 2 below.

Table 2: The table represents the type of restriction, number of parameters within that restriction and number of restrictions associated with different restricted forms of MSAM.

	Type of Restriction	Num. of Parameters	Num. of Restrictions
RC	$\beta_{kji} = 0$ for $j \neq i$	KI	$KI(I-2)$
RE	$\beta_{kji} = \beta_k$	K	$K(I-1)$
RCM	$\Sigma_{ij} = 0$ if $j \neq i$	I	$\frac{1}{2}(I-1)(I-2)-1$

4.2 Estimation

Once we have linearized the market share attraction model, estimating and interpreting the coefficients is pretty straight forward. The linear model can be written as (Franses and Montgomery, 2002):

$$\begin{array}{ccccccc} y_{1t} & = & u'_{1,t}b_1 & + & v'_{1,t}a & + & \eta_{1,t} \\ y_{2,t} & = & u'_{2,t}b_2 & + & v'_{2,t}a & + & \eta_{2,t} \\ \vdots & = & \vdots & + & \vdots & + & \vdots \\ y_{I-1,t} & = & u'_{I-1,t}b_{I-1} & + & v'_{I-1,t}a & + & \eta_{I-1,t} \end{array}$$

where the dependent variable can be denoted as

$$\ln M_{it} - \ln M_{It} = y_{it}$$

$$(\eta_{1t}, ..., \eta_{I-1,t})' = \eta_t \sim N(0, \Sigma^*)$$

u_{it} : these are explanatory variables with brand specific parameter vector b_i.

v_{it} : these are explanatory variables with common parameter vector a.

Due to the potentially large number of equations it is more convenient to rewrite the system in vector-matrix notation. For $i = 1, ..., I-1$ we can define the following vectors and matrices;

$$\begin{aligned} (y_{i1}, ..., y_{iT})' &= y_i \\ (u_{i1}, ..., u_{iT})' &= U_i \\ (v_{i1}, ..., v_{iT})' &= V_i \\ (\eta_{i1}, ..., \eta_{iT})' &= \eta_i \end{aligned}$$

Consequently the problem can be rewritten in matrix notation as:

$$\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_{I-1} \end{pmatrix} = \begin{pmatrix} U_1 & 0 & . & . & . & V_1 \\ 0 & U_2 & . & . & . & V_2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & U_{I-1} & V_{I-1} \end{pmatrix} \begin{pmatrix} b_1 \\ . \\ . \\ b_{I-1} \\ a \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ . \\ . \\ \eta_{I-1} \end{pmatrix}$$

or

$$y = Xb + \eta,$$

where $\eta \sim N(0, (\Sigma^* \otimes I_T))$ and $\otimes$ denotes the Kronecker product.

4.2.1 Market share attraction model (with dynamics)

The fully extended market share attraction model mentioned and studied in the previous section does not include the dynamic effects. Dynamic effects can be defined as the consequences or the outcomes resulted because of the changes done by the market forces (Franses and Montgomery, 2002). These changes can be the altercation in the prices of the goods or the behavioral changes in consumer's consumption or producer's production. The changes or the effects because of these market forces leads to either a temporary change or a permanent change in market. For example, the marketing promotion done in a given time period t might also have its effects in the time $t+1$. This can be defined as a temporary change. In Elberg et al. (2019) the researcher studied the dynamic effects of price promotion in the retail industry in Chile which lead to the discovery of some pretty interesting results within that given market.

Under the dynamic specification of market share attraction model, the attraction of brand i at time t is defined as (Franses and Montgomery, 2002; Erickson, 2002):

$$A_{it} \quad = \quad \exp(\mu_i + \epsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}} \prod_{j=1}^{I} \prod_{p=1}^{P} M_{j,t-p}^{\alpha_{pji}}, \tag{4.7}$$

where α_{pji} shows the effect of p-lagged market share of brand j on attraction of brand i. Here we can see that the attraction of brand i being dependent on the market shares of other brands leads to the generalisation of the dynamic behaviour specification of the market as a whole. Hence, we can see that a change in the market share of other brands certainly have an impact on the attraction as well as the market share of brand i which the researcher will further explore below in the following paragraphs.

Dynamic specifications can further be divided into the following categories depending upon the type of restrictions we use (Franses and Montgomery, 2002; Erickson, 2002):

a) Restricted dynamics: at a given time t, the attraction of brand i only depends upon its own lagged market shared meaning $\alpha_{pji} = 0$, $i \neq j$. This specifications has $PI(I-2)$ restrictions.

b) Common dynamics: in this case, for all brands the dynamics effects is considered to be equal where $\alpha_{pji} = 0$, $i \neq j$ and $\alpha_{pii} = \gamma_p \; \forall i$. This specifications has $P(I-1)$ restrictions.

The specification for the restricted models under dynamics section can be mathematically expressed as follows (Franses and Montgomery, 2002):

a) Restricted dynamics: the respective attraction model and reduced form of market share attraction model are written as follows:

$$
\begin{aligned}
A_{it} &= \exp(\mu_i + \epsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}} \prod_{j=1}^{I} \prod_{p=1}^{P} M_{i,t-p}^{\alpha_{pi}}, \\
\ln M_{it} - \ln M_{It} &= \mu_i^* + \sum_{j=1}^{I} \sum_{k=1}^{K} \beta_{kji}^* \ln x_{kjt} + \sum_{p=1}^{P} (\alpha_{pi} \ln M_{i,t-p} - \alpha_{pI} \ln M_{I,t-p}) + \eta_{it}.
\end{aligned}
$$

b) Common dynamics: the respective attraction model and reduced form of market share attraction model are written as follows:

$$
\begin{aligned}
A_{it} &= \exp(\mu_i + \epsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{kjt}^{\beta_{kji}} \prod_{j=1}^{I} \prod_{p=1}^{P} M_{i,t-p}^{\gamma_p}, \\
\ln M_{it} - \ln M_{It} &= \mu_i^* + \sum_{j=1}^{I} \sum_{k=1}^{K} \beta_{kji}^* (\ln x_{kjt}) + \sum_{p=1}^{P} \gamma_p (\ln M_{i,t-p} - \ln M_{I,t-p}) + \eta_{it}.
\end{aligned}
$$

Interpretation of Market Share attraction model

Working with the fully extended market share attraction model can be quite tiresome and time-consuming. Another problem with the fully extended model is that the interpretation of the parameters is quite troublesome as there are a large number of parameters in this model (Franses and Montgomery, 2002; Erickson, 2002). Also, the presence of dynamics effect makes the interpretation of the parameters tough and the parameters estimated within the full extended model gives the information about the relative market shares. If we want to get the information about the market shares of the brands then we need to work with the reduced form model and then estimate the parameters which can be easily interpreted then.

Simplified model: The equation is written as follows (Franses and Montgomery, 2002):

$$
A_{it} = \exp(\mu_i + \epsilon_{it}) \prod_{j=1}^{I} \prod_{k=1}^{K} P_{jt}^{\beta_{ji}}, \tag{4.8}
$$

Here the partial effect of price on market share is defined as:

$$
\frac{\partial M_{it}}{\partial P_{it}} = \sum_{k=1}^{I} \frac{\partial M_{it}}{\partial A_{kt}} \frac{\partial A_{kt}}{\partial P_{it}},
$$

where

$$
\frac{\partial A_{kt}}{\partial P_{it}} = \frac{\beta_{ik}}{P_{it}} A_{kt}
$$

We know, $\frac{\partial M_{it}}{\partial A_{it}} = \frac{1-M_{it}}{\sum_{j=1}^{I} A_{jt}}$ and $\frac{\partial M_{it}}{\partial A_{kt}} = \frac{-M_{it}}{\sum_{j=1}^{I} A_{jt}}$ for $(k \neq i)$ so that,

$$\frac{\partial M_{it}}{\partial P_{it}} = \frac{M_{it}}{P_{it}}(\beta_{ii} - \sum_{k=1}^{I} \beta_{ik} \, M_{kt})$$

From the above equation the **Price Elasticity** can be shown as follows:

$$\frac{\partial M_{it}}{\partial P_{it}} \frac{P_{it}}{M_{it}} = \beta_{ii} - \sum_{k=1}^{I} \beta_{ik} \, M_{kt} \tag{4.9}$$

From equation 4.9 it is clear that the elasticity goes to zero once the market share approaches to one which seems acceptable in the day to day mechanics of market.

Forecasting relative market shares

We know that the market share attraction model can be written as:

$$\ln M_{it} - \ln M_{It} = \mu_i^* + \sum_{j=1}^{I} \sum_{k=1}^{K} \beta_{kji}^* \, \ln x_{kjt} + \sum_{p=1}^{P} \sum_{j=1}^{I} \alpha_{pji}^* \, \ln M_{j,t-p} + \, \eta_{it}, \tag{4.10}$$

which can be further summarised as:

$$\ln m_{it} = \ln M_{it} - \ln M_{It} = \lambda_{it} + \, \eta_{it}, \tag{4.11}$$

Here $\eta_{it} = (\eta_{1t}, ..., \eta_{I-1,t})' \sim N(0, \Sigma^*)$ which implies that,

$$m_{it} \sim LogN(\lambda_{it}, \Sigma_{ii}^*), \qquad \forall \, i \; at \; time \; t \tag{4.12}$$

Therefore we get,

$$\tilde{m}_{it} = \exp\left(\lambda_{it}, \frac{1}{2} \sum_{ii}^{*}\right), \qquad i = 1, ..., I-1$$

where, $\tilde{m}_{it} = E(m_{it})$

Forecasting market shares

One of the benefits of using market share attraction model is its ability to make good forecasts. A forecast of market share of brand $i = 1, ...I$ at a given time interval $t+1$ is given by:

$$\mathrm{E}[M_{it} | D_{t+1}, \, M_t] \; ,$$

where

D_{t+1}: Information about explanatory variables till time period $t+1$.

M_t: Information about market shares till time period t.

In order to calculate the market shares we need to write M_{it} in the form of an equation relating it to m_{it}. The forecasts of M_{it} can therefore be written as:

$$\tilde{M}_{I,t+1} = E\left[\frac{m_{i,t+1}}{1+\sum_j^{I-1} m_{j,t+1}} | D_{t+1}, M_t\right], \quad i = 1, ..., I-1$$

$$\tilde{M}_{I,t+1} = E\left[\frac{1}{1+\sum_j^{I-1} m_{j,t+1}} | D_{t+1}, M_t\right]$$

But the expected values of m_{it} used in the above equations do not give unbiased forecasts because of the following reason (Franses and Montgomery, 2002):

$$E\left[\frac{X}{Y}\right] \neq \frac{E[X]}{E[Y]} \tag{4.13}$$

In order to get the unbiased forecasts we can consider calculating forecasts via the method of simulations but this falls outside the scope of the current research study.

4.3 Static panel data models

Before going into further detail about the different static panel data models used in this research, we start with a short overview of panel data. Panel data set can be defined as a repeated collection of observations for a given number of individuals N at different points in time T (Wooldridge, 2016).

There are several advantages to using panel data. First, thanks to having both time and spatial dimension they usually consist of a large number of observations which leads to a precise estimation of parameters. This is so because in many parametric models degrees of freedom and power of most statistical tests generally increases with the number of observations (Pesaran, 2015; Hsiao, 2007). Furthermore, having a data set with data and variation in both time and cross-sectional dimensions allows for explicit modeling of unobserved individual specific and time invariant heterogeneity even in the case of the variables behind this are not observable. Thus an advantage of panel data is that it also allows us to control for, or at least reduce, omitted variable bias (Pesaran, 2015; Hsiao, 2007).

The recent increase in panel data availability also led to development of wide plethora of models that can be used to analyze the panel data (See the overview of different methods in Pesaran, 2015). However, this work is only going to use two of these, namely the pooled model and random effects model. This is because, as the rest of this chapter explains, these models are appropriate for the analysis done in this research.

4.3.1 Pooled model

The pooled model essentially treats the panel data as a large cross-section data set which is analyzed by OLS (Pesaran, 2015; Wooldridge, 2016). Hence the pooled model can be specified as follows:

$$y_{it} = \alpha + x'_{it}\beta + \epsilon_{it}, \tag{4.14}$$

where y_{it} is the $1 \times k$ vector of outcome recorded for observation of the i^{th} cross-sectional unit at time t, for $i = 1, 2, ..., N; t = 1, 2, ..., T$, α is a homogeneous intercept for all i, x'_{it} is $1 \times k$ vector of the independent observations, β is $k \times 1$ vector of estimated coefficients and finally ϵ_{it} is the *iid* error (Pesaran, 2015).

Furthermore, according to Pesaran (2015) important assumptions to assure that pooled OLS is unbiased are that x_{it} is strictly exogenous and the intercept α is indeed homogeneous. As standard OLS, the estimates are unbiased but inconsistent in the presence of heteroskedasticity or autocorrelation. In addition in panel data consistency also depends on assumption that errors are cross-sectionally independent.

As argued by Wooldridge (2016) the main advantages of using pooled model is the efficiency and parsimony of the estimator compared to other panel data models. Poled model is computationally very simple to use and it requires less predictors compared to alternatives. However, its simplicity comes at a 'cost' of several drawbacks. First drawback of using this model is that it does not allow for heterogeneity across cross-section members of the panel, as all intercepts are assumed to be homogenous, which is not realistic in most of the real life based marketing case scenarios. This as mentioned above may violate the model assumptions and lead to bias. Secondly, pooled model estimates are inconsistent if fixed effects is true data generating process (DGP). Both of these drawbacks are visualised on the figure 1 below.

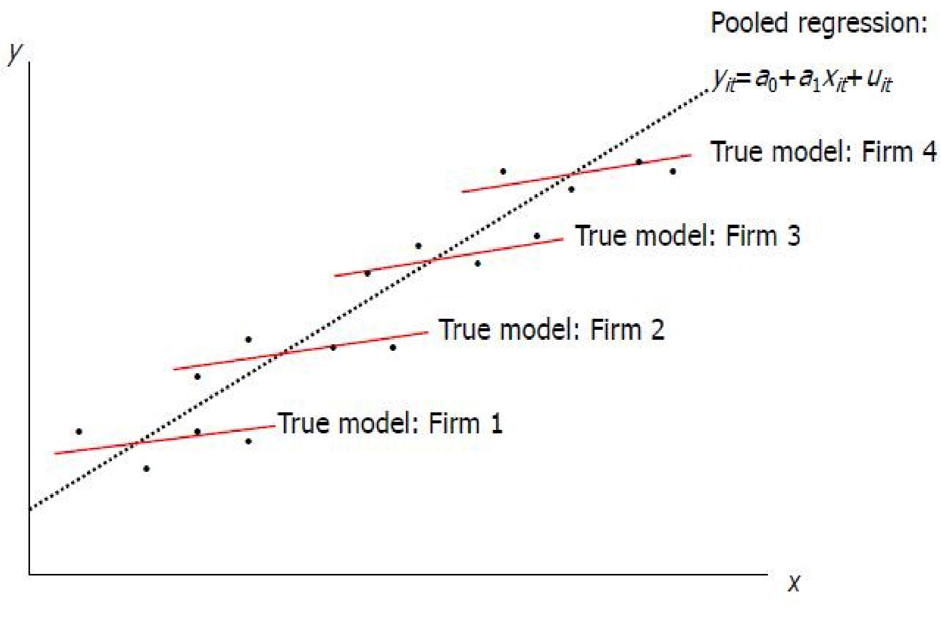


Figure 1: Pooled regression

4.3.2 Individual-specific effects model (Random effects)

In contrast to the pooled model, random effects model relaxes the rather strict and unrealistic assumption that $\alpha = \alpha_i, \forall i$. Random effects specification istead assumes that α_i are realizations from a pre-specified probability distribution function (usually Gaussian distribution) assuming that α_i is uncorrelated with the explanatory variables x_{it} (Pesaran, 2015). Thus, the effects of α_i can be coupled with the error term. The model is specified as follows:

$$y_{it} \;=\; \alpha \;+\; x'_{it}\beta \;+\; u_{it} \tag{4.15}$$

where

$u_{it} \;=\; (\alpha_i - \alpha) \;+\; \epsilon_{it}$,

α is the constant effect,

$\alpha_i \sim iid\ (\alpha, \sigma^2_\alpha)$ is the individual effect assumed to be uncorrelated with x_{it},

$\epsilon_{it} \sim iid\ (0, \sigma^2_\epsilon)$ is the disturbance/noise,

Random effects model can be consistently estimated by pooled estimators, between estimators and feasible generalised least square estimators. For this particular case study we use the between estimators criteria for estimation purposes. In Dieleman and Templin (2014) the researcher showcases the advantage of implementing the random effects model with a given dataset using the between estimators criteria despite the estimators showing some biasedness. Random effects' estimation were found to be more precise under certain conditions than the competing estimators in the above mentioned research study.

Between estimators works on the criteria of using the cross-sectional variations for parameter estimation as opposed to time variations. The original individual-specific effects model is specified as:

$$y_{it} = \alpha_i \;+\; x'_{it}\beta \;+\; \epsilon_{it} \tag{4.16}$$

$$= \alpha \;+\; x'_{it}\beta \;+\; (\alpha_i - \alpha + \epsilon_{it}) \tag{4.17}$$

Taking the averages on both sides w.r.t to time,

$$\frac{1}{T}\sum_{t=1}^{T} y_{it} \;=\; \frac{1}{T}\sum_{t=1}^{T}(\alpha \;+\; x'_{it}\beta \;+\; \alpha_i - \alpha + \epsilon_{it}) \tag{4.18}$$

which comes out to be,

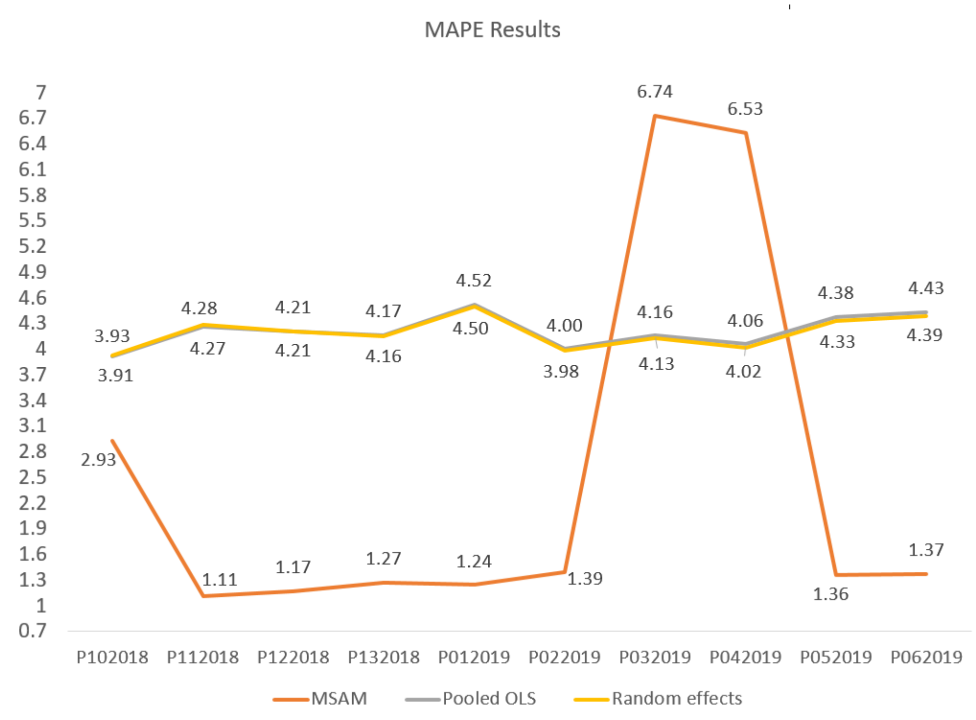
$$\bar{y}_i \;=\; \alpha_i \;+\; \bar{x}'_i\beta \;+\; \bar{u}_i. \tag{4.19}$$

From the above equations we can see that the between estimator is a basic OLS regression of $\bar{y}_i$ on intercept and $\bar{x}_i$ thus making use of the variations in the dependent variable and explanatory variables between individuals.

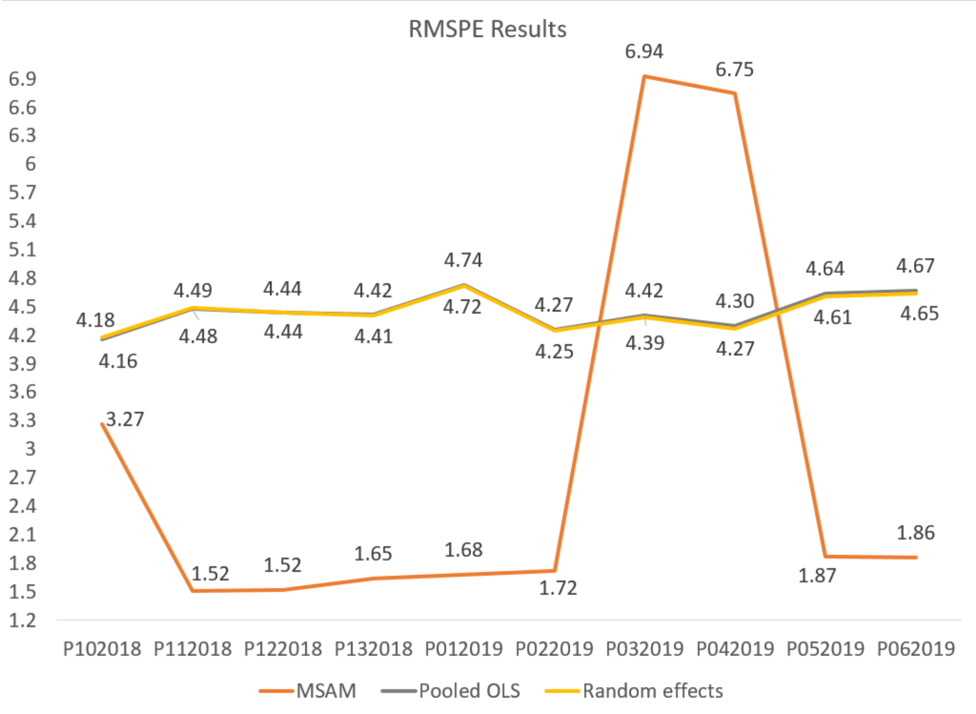
One of the drawbacks of between estimators is that it is only consistent when the pooled or random effects model is the true DGP. Between estimators are inconsistent in case of a correlation between α_i and x_{it}. Furthermore, as in the previous case heteroskedasticity, autocorrelation or cross sectional dependence could lead to inconsistent estimation.

5 Results

The researcher has implemented pooled OLS, random effects and MSAM OLS methods for this research study and the results are presented below. MSAM is an extremely powerful and advanced model used immensely to deal with the limitations related with individual effects and problems concerned with those model where the lag values of explanatory variables are not included. In figures 2a and 2b, the mean average prediction error (MAPE) and root mean square prediction error (RMSPE) are presented respectively in order to evaluate the performance of the models. In the figures 2a and 2b, clearly we can see that MSAM OLS performs the best in terms of both MAPE and RMSPE for majority of time periods. In the same sample time periods, both pooled OLS and random effects performs the worst as they show approximately three times the value of MSAM's MAPE and RMSPE results as shown in the figures. Both pooled OLS and random effects have approximately the same values for both MAPE and RMSPE results, showing that the inclusion of individual heterogeneity in the random effects model did not lead to significant change in the prediction values from those presented within the pooled OLS model. In figures 2a and 2b, there are about 2 time periods where MSAM OLS performs the worst in terms of both MAPE and RMSPE results. But for better part of the out of sample periods, MSAM OLS is able to overcome the adverse effects of the bad performing days of the market. Also, a noteworthy point to point out here is that, a lot of new product range within the fabric solutions category were introduced in the market during the periods P032019 and P042019. Products were launched not only under the fabric solutions category of Unilever but other companies like Henkel and Colgate-Palmolive were also heavily invested in the launch of new products and their different variants. The launch of Eco-green products within different markets of EU is one such example of the new product launches in the time period P032019 and P042019. These products, because of their lack in the lag explanatory values of the data, calculated some really biased predictions due to which the MAPE and Results for these periods within MSAM, shoots to a really high value as seen in the figures 2a and 2b. However, the predictions results don't get affected within the pooled OLS and random effects model because of the lack of lag variables in them and hence the MAPE and RMSPE values of periods P032019 and P042019, within these two models are consistent with the values of the previous time-periods.



(a) Average MAPEs over all brands for pooled OLS, random effects, Average RMPSE's over all brands for pooled OLS, random effects, MSAM OLS for out of sample 10 periods



(b) Average RMPSE's over all brands for pooled OLS, random effects, MSAM OLS for out of sample 10 periods

Table 3: Regression results of OLS, Random effects

	Ordinary least squares		*Random effects*	
	Estimates	Pr(> $\|t\|$)	Estimates	Pr(> $\|t\|$)
Constant	−11.522	0.000	−10.900	0.000
***	1.640		1.918	
Volume sales	0.704	0.000	0.658	0.000
***	0.097		0.115	
Promo sales	0.024	0.001	0.019	0.016
***	0.036		0.041	
Care	0.107	0.081	0.120	0.084
***	0.040		0.048	
Eco	0.588	0.000	0.624	0.000
***	0.108		0.130	
Frag fresh	0.144	0.024	0.143	0.039
***	0.096		0.118	
Frag perfume	0.084	0.312	0.084	0.382
***	0.092		0.113	
Hygiene	0.094	0.786	0.073	0.853
***	0.456		0.537	
Sensitive	0.353	0.000	0.383	0.000
***	0.073		0.087	
Stain removal	0.319	0.000	0.334	0.000
***	0.052		0.062	
Whiteness	0.192	0.054	0.199	0.084
***	0.062		0.072	
Color	0.012	0.879	0.009	0.919
***	0.031		0.036	
Baby	−0.374	0.106	−0.429	0.107
***	0.139		0.164	

[*] *Note:* The dependent variable is the MS of products with respect to competitor. The table reports the coefficients estimated via OLS and random effects on entire sample of 705 fabric solutions products across 38 time periods ranging from 2016 till 2019. The explanatory variables includes promo sales, volume sales and market share of other products. The p-value of the estimates are mentioned in the table as well. The rows with *** represents the std errors of the estimators.

Table 4: **Summary of pooled OLS, random effects and MSAM OLS**

	OLS	Random effects	MSAM OLS
SSR	228070	225200	20125.44
R-Squared	0.1506	0.1132	0.5396
Adj. R-Squared	0.1503	0.1130	0.4704
Chisq	–	3333.27 on 12 DF	–
F-statistic	85.382 on 12 & 26072 DF	–	7.797 on 92 & 612 DF

* *Note:* Table 4 represents the summary statistics of the pooled OLS, random effects and OLS for MSAM comprising of R-squared values, sum of squared residuals (SSR), Chisquare and F-statistics.

The Table 4 presents the summary of pooled OLS, random effects and MSAM OLS respectively. It shows how well the MSAM performs with respect to estimating the parameters of the model and gives the best goodness of fit value (R^2) along with the lowest value for sum squared of residuals within the three models that the researcher has used for this research study.

Table 5: **MAPE & RMPSE results**

	Pooled OLS	RE	MSAM	IMP of MSAM on Pooled	IMP of MSAM on RE
MAPE	4.2112	4.1941	2.5119	1.676 times	1.669 times
RMPSE	4.4529	4.4419	2.8766	1.547 times	1.544 times

* *Note:* Table 5 shows the average MAPE and RMPSE results of 10 out of sample periods ranging from P10 2018 till P06 2019 for OLS, random effects and MSAM OLS. These results are computed for the entire 705 products for the out of sample period. "IMP" in the above table is the abbreviation which means improvement. RE stands for random effects.

In Table 5, the researcher presents the average periods MAPE and RMSPE values for pooled OLS, random effects and MSAM OLS for out of sample 10 periods. The table showcases how well the MSAM OLS outperforms the other methods. In MSAM, the lag variables and individuals effects are introduced and implemented in an efficient way which leads to a higher goodness of fit (R^2) value as compared to the other two models and also helps in getting the unbiased estimates of the coefficients which appears to be biased in both pooled OLS and random effects model because of the missing lag explanatory variables. Both the above mentioned points helps in getting better predicted results from MSAM as compared to the other two models and that is precisely the reason why the average periods MAPE and RMSPE values for a out of sample of 10 periods is lowest for the MSAM and is atleast 1.5 times better than the average MAPE and RMSPE results of pooled OLS and random effects.

All the results are obtained while working on 'Windows 10 Enterprise' on the system 'Lenovo' model name ‘ThinkPad T480’ with 64-bit operating system. The processor specification for the system is ‘Intel(R) core(TM) i5-8350U CPU @ 1.70GHz 1.90GHz’. The software platform used for this research purpose is ‘R x64 3.5.2’ with the code written in the programming language called ‘R’.

6 Conclusions

As seen from the results presented under the Section 5, it is recommended to use MSAM while calculating the relative market share of the products because it produces unbiased estimates of the parameters within our given dataset and includes lag explanatory variables within the working of model which gives much more accurate predictions as compared to both pooled OLS and random effects. One of the drawbacks of using market share attraction model with the available data-set is that it endures a lot of estimation problems if the assumptions within the model are not satisfied. If there are problems in the model while working, one cannot expect accurate predictions to be obtained via the model. For example, the biased prediction results which lead to high MAPE and RMSPE results within MSAM for periods P032019 and P042019. An elegant and intuitive way to counter this drawback is to include the use of non-parametric methods such as support vector regressions with the present data-set but this falls outside the scope of our present research study.

Thus, in the end, i can say that there is ample evidence in my research study that MSAM performs the best. Although there might be a possibility of some discrepancies while doing predictions with MSAM as seen in the results but still for better part of our prediction results, MSAM outbid both pooled OLS and random effects in every department of statistical study.

7 About Software and programming language

In this research study, the researcher has used the ‘R-studios (R x64)’ platform version ‘3.5.2’ to work with the given data set. The programming language used in this is ‘R’. ‘R’ is a widely renowned and used programming language for statistical and analytical purposes and is supported by the R Foundation for Statistical Computing.

Some of the important packages used in this research study are as follows. In order to read, write and format excel files the researcher has used ‘xlsx’ package. ‘dplyr’ is a package that is implemented within this research study in order to manipulate datasets more efficiently in R. ‘dplyr’ is the next iteration of ‘plyr’, focusing on only on data frames. With the help of ‘dplyr’, the researcher is able to separate the data into test and training sets during the implementation of pooled OLS, random effects and MSAM. ‘plm’ package is used to run linear regression and estimation of the variables in the given model for panel data sets. ‘prediction’ package is used to perform the prediction of relative market shares from the given data set using the market share attraction model. ’ggplot2’ package is used to plot graphs and figures that the researcher presents in this research study related with the data and the outputs. ‘MLmetrics’ package is used for the evaluation purpose while running regression with the given data set.

8 Discussion on future scope

There is a lot of scope of discussion and implementation of various functionalities with the current data-set. First of all, the effect of different combinations of benefits indicators can be used in order to see how it effects the relative market share of products and how it effects the purchasing behaviour of consumers. A benchmark product having different numbers of benefits associated with it can be used randomly to see how it determines the relative market share. Also, instead of relative market shares, the market shares of products can be calculated by making use of simulation schema.

The implementation of non-parametric methods such as support vector regressions (SVR) can also be done with the given models and the current data-set. In regards to the SVR implementation, we can use both the linear and non-linear approaches. Under the non-linear approach, a further investigation on choosing an efficient procedure to tune in manually adjustable parameter in SVR can be done as well. We can even look forward for different ways in order to compute confidence intervals for predictions and also how to interpret the parameters for non-linear relationship of the model. Different approaches can further be investigated in order to accept or deny with conviction the reliability and usefulness of SVR method with respect to market share attraction model and marketing related data-sets.

References

Balestra, P. and Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica (pre-1986)*, 34(3):585.

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of forecasting*, 27(2):153–173.

Bell, D. E., Keeney, R. L., and Little, J. D. (1975). A market share theorem. *Journal of Marketing Research*, 12(2):136–141.

Berkey, C. S., Hoaglin, D. C., Mosteller, F., and Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in medicine*, 14(4):395–411.

Blossfeld, H.-P., Schneider, T., and Doll, J. (2009). Methodological advantages of panel studies: Designing the new national educational panel study (neps) in germany. *Journal for Educational Research Online/Journal für Bildungsforschung Online*, 1(1):10–32.

Bultez, A. (1978). Econometric specification and estimation of market share models: The state of the art. In *Marketing*, pages 239–263. Springer.

Calver, A. (2001). Riverbed permeabilities: Information from pooled data. *Groundwater*, 39(4):546–553.

Dieleman, J. L. and Templin, T. (2014). Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: a simulation study. *PloS one*, 9(10):e110257.

Dingemanse, N. J. and Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, 82(1):39–54.

Elberg, A., Gardete, P. M., Macera, R., and Noton, C. (2019). Dynamic effects of price promotions: Field evidence, consumer search, and supply-side implications. *Quantitative Marketing and Economics*, 17(1):1–58.

Erickson, G. (2002). *Dynamic models of advertising competition*, volume 13. Springer Science & Business Media.

Fianto, A. Y. A., Hadiwidjojo, D., Aisjah, S., et al. (2014). The influence of brand image on purchase behaviour through brand trust. *Business Management and Strategy*, 5(2):58.

Fieuws, S. and Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statistics in medicine*, 23(20):3093–3104.

Fok, D. and Franses, P. H. (2004). Analyzing the effects of a brand introduction on competitive structure using a market share attraction model. *International Journal of Research in Marketing*, 21(2):159–177.

Franses, P. H. and Montgomery, A. L. (2002). *Econometric models in marketing*, volume 16. Elsevier.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of econometrics*, 126(2):269–303.

Hanssens, D. M., Parsons, L. J., and Schultz, R. L. (2003). *Market response models: Econometric and time series analysis*, volume 12. Springer Science & Business Media.

Hogeweg, P. and Hesper, B. (1990). Individual-oriented modelling in ecology. *Mathematical and Computer Modelling*, 13(6):83–90.

Hsiao, C. (2007). Panel data analysisadvantages and challenges. *Test*, 16(1):1–22.

Karnani, A. (1985). Strategic implications of market share attraction models. *Management Science*, 31(5):536–547.

Kittel, B. (1999). Sense and sensitivity in pooled analysis of political data. *European Journal of Political Research*, 35(2):225–253.

Kopalle, P. K., Mela, C. F., and Marsh, L. (1999). The dynamic effect of discounting on sales: Empirical analysis and normative pricing implications. *Marketing Science*, 18(3):317–332.

Kotler, P. (1965). Competitive strategies for new product marketing over the life cycle. *Management Science*, 12(4):B–104.

Kotler, P. and Armstrong, G. (2010). Principles of marketing: Upper saddle river.

Ma, R., Krewski, D., and Burnett, R. T. (2003). Random effects cox models: a poisson modelling approach. *Biometrika*, 90(1):157–169.

Monahan, G. E. and Sobel, M. J. (1994). Stochastic dynamic market share attraction games. *Games and Economic Behavior*, 6(1):130–149.

Naert, P. and Weverbergh, M. (1981). On the prediction power of market share attraction models. *Journal of Marketing Research*, 18(2):146–153.

Nakanishi, M. and Cooper, L. G. (1974). Parameter estimation for a multiplicative competitive interaction modelleast squares approach. *Journal of marketing research*, 11(3):303–311.

Nakanishi, M. and Cooper, L. G. (1988). Market share analysis: Evaluating competitive marketing effectiveness.

Pauwels, K., Hanssens, D. M., and Siddarth, S. (2002). The long-term effects of price promotions on category incidence, brand choice, and purchase quantity. *Journal of marketing research*, 39(4):421–439.

Pesaran, M. H. (2015). *Time series and panel data econometrics*. Oxford University Press.

Setyawan, A. A. and Imronudin, K. (2015). Brand trust and brand loyalty, an empirical study in indonesia consumers. *British Journal of Marketing Studies*, 4(3):37–47.

Simon, H. (1978). An analytical investigation of kotler's competitive simulation model. *Management Science*, 24(14):1462–1473.

Tinbergen, J. (1940). Econometric business cycle research. *The Review of Economic Studies*, 7(2):73–90.

Victoria, A. C. and Ganesan, M. (2014). Online sales promotions of grocery and other fmcg products in chennai entity. *Middle-East Journal of Scientific Research*, 22(10):1520–1526.

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

Appendix I: Code

This appendix showcases the code that was used to derive all results included in this work. The full list of packages used while writing this code includes: xlsx; fastDummies; dplyr; plm; stargazer; lmtest; readxl; robustbase; xts; e1071; ggplot2; Matrix; fpp2; ROCR; prediction; Metrics; MLmetrics; useful; data.table; DataCombine.

Data Import, Cleaning and Summary Statistics

The first step in this data analysis was importing, cleaning and adjusting the dataset, which was done using the following code:

```
mydata<-read_excel("Finalised Data file.xlsx",
                        sheet = "Raw Data")
mydataa<- data.frame(mydata)
head(mydataa)
vvv <-summary(mydataa)
usethrow<-dummy_cols(mydataa[,85:94])
usethrow1<-usethrow[, c(11, 14, 16, 18, 20, 22, 24, 26, 28,
                         30)]
mydataa[,c(85:94)]<-usethrow1[,c(1:10)]
mydataa[is.na(mydataa)] <- 0
```

Next, the summary statistics, mainly of the KPI variables (see chapter 3), for the sample used in this study was estimated using the code below

```
bu<-as.data.frame(colnames(mydataa))
barp<-function(hi){
    for(i in 1:132){
      hi<-mydataa[,i]
  par(mar=c(5,8,4,2))
  barplot(prop.table(table(hi)), xlab = bu[i,])
    }
}
par(mfrow = c(4,5))
barp(hi)
```

Panel Data

The final panel dataset is derived with the code shown below. To ensure replicability random seed is specified before proceeding.

```
set.seed(99)
```

```r
 jug<-data.frame(mydataa[,95:132])
 benchmark<-data.frame()
 benchmark_1 <-data.frame()
    for(i in 1:38){
      for(j in 1:nrow(jug))
           benchmark[j,i]<-log(jug[303,i])
    }

ki<-list(mydataa[,c(9:46)], mydataa[,c(47:84)],
           mydataa[,c(95:132)], benchmark[,c(1:38)] )

result_fc<- list()
for (fc in 1:4) {
  cool_panel1_fc<-c()
  for (v in 1:nrow(ki[[fc]])){
    small_fc=t(ki[[fc]][v,1:38])
    cool_panel1_fc=rbind(cool_panel1_fc,small_fc)
  }
  result_fc[[fc]]<-cool_panel1_fc
}

adi<-list(mydataa[,c(1)],mydataa[,c(2)],mydataa[,c(3)],
            mydataa[,c(4)],mydataa[,c(5)],mydataa[,c(6)],
            mydataa[,c(7)],mydataa[,c(8)],mydataa[,c(85)],
            mydataa[,c(86)],mydataa[,c(87)], mydataa[,c(88)],
            mydataa[,c(89)],mydataa[,c(90)], mydataa[,c(91)],
            mydataa[,c(92)], mydataa[,c(93)],mydataa[,c(94)])
result_fcm<- list()
for (fcm in 1:18) {
  notcool_panel1_fcm<-c()
  for (vm in 1:length(adi[[fcm]])){
    large_fcm=rep(adi[[fcm]][vm],38)
    notcool_panel1_fcm=append(notcool_panel1_fcm,large_fcm)
  }
  result_fcm[[fcm]]<-notcool_panel1_fcm
}

paneldata123<-as.data.frame(matrix(unlist(result_fc),
             byrow = F, nrow=26790))
```

```r
names(paneldata123)<- c("Promo_Sales","Volume_Sales",
"Market_Share", "Benchmark")

paneldata1234<-as.data.frame(matrix(unlist(result_fcm),
byrow = F, nrow=26790))
names(paneldata1234)<- c("Country", "Company", "Global_brand",
                         "Brand", "Segment", "Format",
                         "Variant", "Actual_Pack_Size", "Care",
                         "Eco","Frag_fresh","Frag_perfume",
                         "Hygiene","Sensitive","Stain_Removal",
                         "Whiteness", "Color", "Baby")
MAIN_PANEL_DATA<-cbind(paneldata1234, paneldata123)

products_id_panel1<-c()
for (j in 1:nrow(mydataa)){
  b=rep(j,38)
  products_id_panel1=append(products_id_panel1,b)
}
products_id_panel1<-as.data.frame(products_id_panel1)
names(products_id_panel1)<- c("Product_ID")
years_panel1=as.data.frame(rep(seq(1,38, by = 1),
times = nrow(mydataa), each = 1))
names(years_panel1)<- c("PERIODS")
MAIN_PANEL_DATA<-cbind(products_id_panel1,years_panel1,
                       MAIN_PANEL_DATA )f

FINALIZED_PANEL<-pdata.frame(MAIN_PANEL_DATA ,
                             index=c("Product_ID", "PERIODS"))
l <-summary(FINALIZED_PANEL)
FINALIZED_PANEL$Market_Share<-log(FINALIZED_PANEL$Market_Share)
FINALIZED_PANEL$Volume_Sales<-log(FINALIZED_PANEL$Volume_Sales)
FINALIZED_PANEL$Promo_Sales<-log(FINALIZED_PANEL$Promo_Sales)

FINALIZED_PANEL[, c(11:20)]<-lapply(FINALIZED_PANEL[,
                             c(11:20)], FUN= as.character)
FINALIZED_PANEL[, c(11:20)]<-lapply(FINALIZED_PANEL[,
                             c(11:20)], FUN= as.numeric)
```

Pooled OLS and Random Effects Models

This subsection finally presents the code used to estimate the pooled OLS and Random effects models (see the section 4 for more details on estimation procedure). Furthermore, the code at the end collects the coefficients for predictions from both models.

```
c<-list()
d<-list()
u<-list()
v<-list()
y<-list()
z<-list()
N<-28
nn<-0
for (i in seq(N)) {
  while(nn < 10){
  a<-  dplyr::filter(FINALIZED_PANEL,
                          PERIODS %in% c(1:(N+nn)))
  b<-  dplyr::filter(FINALIZED_PANEL,
                          PERIODS %in% c(2:(N+nn+1)))

  p <- plm(a$Market_Share - a$Benchmark  ~ a$Volume_Sales +
         a$Promo_Sales  + a$Care + a$Eco + a$Frag_fresh +
         a$Frag_perfume  + a$Hygiene + a$Sensitive +
         a$Stain_Removal + a$Whiteness + a$Color + a$Baby
         , data = a, effect = "twoways",  model = "pooling")
  r <- plm(a$Market_Share - a$Benchmark  ~ a$Volume_Sales +
             a$Promo_Sales  + a$Care + a$Eco + a$Frag_fresh +
             a$Frag_perfume  + a$Hygiene + a$Sensitive +
             a$Stain_Removal + a$Whiteness + a$Color + a$Baby,
             data = a, model = "random",
             inst.method =  "baltagi")

   e<- predict(p, b)
   g<- predict(r, b)
```

Afterwards, the predictions are derived for both models using the following simultaneous loop. The code itself is actually not written as a loop statement but written out in full. Consequently, only the beginning and end of the code is presented with the excluded code indicated by ellipsis (...). This was done in order to reduce the code size and make it presentable.

```
e<-as.data.frame(e)
  e<-e[c(1*(N+nn), 2*(N+nn), 3*(N+nn), 4*(N+nn), 5*(N+nn),
       6*(N+nn), 7*(N+nn), 8*(N+nn), 9*(N+nn), 10*(N+nn),
       11*(N+nn), 12*(N+nn), 13*(N+nn), 14*(N+nn), 15*(N+nn),
                              ...
       694*(N+nn), 695*(N+nn), 696*(N+nn), 697*(N+nn),
       698*(N+nn), 699*(N+nn), 700*(N+nn), 701*(N+nn),
       702*(N+nn), 703*(N+nn), 704*(N+nn), 705*(N+nn)),]

  e<-as.data.frame(e)
  names(e)[1]<-"pre"

  g<-as.data.frame(g)
  g<-g[c(1*(N+nn), 2*(N+nn), 3*(N+nn), 4*(N+nn), 5*(N+nn),
  6*(N+nn), 7*(N+nn), 8*(N+nn), 9*(N+nn), 10*(N+nn),
  11*(N+nn), 12*(N+nn), 13*(N+nn), 14*(N+nn), 15*(N+nn),
                              ...
  694*(N+nn), 695*(N+nn), 696*(N+nn), 697*(N+nn),
  698*(N+nn), 699*(N+nn), 700*(N+nn), 701*(N+nn),
  702*(N+nn), 703*(N+nn), 704*(N+nn), 705*(N+nn)),]

  g<-as.data.frame(g)
   names(g)[1]<-"pre"
```

Finally, the forecasting accuracy of these models is examined using the MAE and RMSE (see the results in section 5). Lastly the results are also exported into latex using the stargazer package.

```
   bald<-  dplyr::filter(b, PERIODS %in% c(N+nn+1))
   balder<-bald$Market_Share - bald$Benchmark
   balder<-as.data.frame(balder)
   names(balder)[1]<-"real"

   c[[N+nn]]<-MAE(e$pre, balder$real)
   d[[N+nn]]<-RMSE(e$pre, balder$real)
   y[[N+nn]]<-MAE(g$pre, balder$real)
   z[[N+nn]]<-RMSE(g$pre, balder$real)
   nn<-nn+1
```

```
kc<-unlist(c)
kc<-data.frame(kc[1:10])

kd<-unlist(d)
kd<-data.frame(kd[1:10])

ky<-unlist(y)
ky<-data.frame(ky[1:10])

kz<-unlist(z)
kz<-data.frame(kz[1:10])

p<- coeftest(p, vcov.=vcovSCC)
r<- coeftest(r, vcov.=vcovSCC)

stargazer(p,r, type = "latex")
```

Market share attraction model

This subsection presents the code used to estimate the parameters of MSAM (see the section 4 for more details on estimation procedure). Firstly we start by new variables as shown below which is the relative market share of products from 27^{th} period till the 38^{th} period.

```
mydataa["Y_i27"]<- mydataa[,121]- rep(mydataa[303,121])
mydataa["Y_i28"]<- mydataa[,122]- rep(mydataa[303,122])
mydataa["Y_i29"]<- mydataa[,123]- rep(mydataa[303,123])
mydataa["Y_i30"]<- mydataa[,124]- rep(mydataa[303,124])
mydataa["Y_i31"]<- mydataa[,125]- rep(mydataa[303,125])
mydataa["Y_i32"]<- mydataa[,126]- rep(mydataa[303,126])
mydataa["Y_i33"]<- mydataa[,127]- rep(mydataa[303,127])
mydataa["Y_i34"]<- mydataa[,128]- rep(mydataa[303,128])
mydataa["Y_i35"]<- mydataa[,139]- rep(mydataa[303,129])
mydataa["Y_i36"]<- mydataa[,130]- rep(mydataa[303,130])
mydataa["Y_i37"]<- mydataa[,131]- rep(mydataa[303,131])
mydataa["Y_i38"]<- mydataa[,132]- rep(mydataa[303,132])

NEW_DATA <-as.data.frame(mydataa[, c(133)])
NEW_DATA[, c(2:39)]<-as.data.frame(mydataa[, c(95:132)])
NEW_DATA[, c(40:115)]<-as.data.frame(mydataa[, c(9:84)])
NEW_DATA[, c(116:125)]<-as.data.frame(mydataa[, c(85:94)])
NEW_DATA[, c(126:137)]<-as.data.frame(mydataa[, c(134:145)])
```

Here we start by making some vectors which would be helpful when we run the regression model from 27^{th} period till the 37^{th} period. Also we create a list in order to store the predicted values generated from the MSAM.

```
i<-126
nn<-0
ii<-1
LLLL <- c(0,1,1,1,1,1,1,1,1,1,1,1)
LLLL <-as.data.frame(LLLL)
MMMM <- c(0,0,1,1,1,1,1,1,1,1,1,1)
MMMM <-as.data.frame(MMMM)
NNNN <- c(0,0,0,1,1,1,1,1,1,1,1,1)
NNNN <-as.data.frame(NNNN)
OOOO <- c(0,0,0,0,1,1,1,1,1,1,1,1)
OOOO <-as.data.frame(OOOO)
PPPP <- c(0,0,0,0,0,1,1,1,1,1,1,1)
PPPP <-as.data.frame(PPPP)
QQQQ <- c(0,0,0,0,0,0,1,1,1,1,1,1)
QQQQ <-as.data.frame(QQQQ)
RRRR <- c(0,0,0,0,0,0,0,1,1,1,1,1)
RRRR <-as.data.frame(RRRR)
SSSS <- c(0,0,0,0,0,0,0,0,1,1,1,1)
SSSS <-as.data.frame(SSSS)
TTTT <- c(0,0,0,0,0,0,0,0,0,1,1,1)
TTTT <-as.data.frame(TTTT)
UUUU <- c(0,0,0,0,0,0,0,0,0,0,1,1)
UUUU <-as.data.frame(UUUU)

KAL <-list()
```

He we start with the while loop in order to run the regression for 11 iterations. We keep on increasing the lag in each iteration step.

```
  while(nn<11) {
  NEW_DATA[,29]<-NEW_DATA[,29]* LLLL[ii, ]
  NEW_DATA[,30]<-NEW_DATA[,30]* MMMM[ii, ]
  NEW_DATA[,31]<-NEW_DATA[,31]* NNNN[ii, ]
  NEW_DATA[,32]<-NEW_DATA[,32]* OOOO[ii, ]
  NEW_DATA[,33]<-NEW_DATA[,33]* PPPP[ii, ]
  NEW_DATA[,34]<-NEW_DATA[,34]* QQQQ[ii, ]
  NEW_DATA[,35]<-NEW_DATA[,35]* RRRR[ii, ]
```

```
 NEW_DATA[,36]<-NEW_DATA[,36]* SSSS[ii, ]
 NEW_DATA[,37]<-NEW_DATA[,37]* TTTT[ii, ]
 NEW_DATA[,38]<-NEW_DATA[,38]* UUUU[ii, ]
 NEW_DATA[,67]<-NEW_DATA[,68]* LLLL[ii, ]
 NEW_DATA[,68]<-NEW_DATA[,68]* MMMM[ii, ]
 NEW_DATA[,69]<-NEW_DATA[,69]* NNNN[ii, ]
 NEW_DATA[,70]<-NEW_DATA[,70]* OOOO[ii, ]
 NEW_DATA[,71]<-NEW_DATA[,71]* PPPP[ii, ]
 NEW_DATA[,72]<-NEW_DATA[,72]* QQQQ[ii, ]
 NEW_DATA[,73]<-NEW_DATA[,73]* RRRR[ii, ]
 NEW_DATA[,74]<-NEW_DATA[,74]* SSSS[ii, ]
 NEW_DATA[,75]<-NEW_DATA[,75]* TTTT[ii, ]
 NEW_DATA[,76]<-NEW_DATA[,76]* UUUU[ii, ]
 NEW_DATA[,105]<-NEW_DATA[,105]* LLLL[ii, ]
 NEW_DATA[,106]<-NEW_DATA[,106]* MMMM[ii, ]
 NEW_DATA[,107]<-NEW_DATA[,107]* NNNN[ii, ]
 NEW_DATA[,108]<-NEW_DATA[,108]* OOOO[ii, ]
 NEW_DATA[,109]<-NEW_DATA[,109]* PPPP[ii, ]
 NEW_DATA[,110]<-NEW_DATA[,110]* QQQQ[ii, ]
 NEW_DATA[,111]<-NEW_DATA[,111]* RRRR[ii, ]
 NEW_DATA[,112]<-NEW_DATA[,112]* SSSS[ii, ]
 NEW_DATA[,113]<-NEW_DATA[,113]* TTTT[ii, ]
 NEW_DATA[,114]<-NEW_DATA[,114]* UUUU[ii, ]

     MSAM <-lm(NEW_DATA[,i] ~ NEW_DATA[,1]+ NEW_DATA[,2] +
     NEW_DATA[,3] +NEW_DATA[,4]+ NEW_DATA[,5]+ NEW_DATA[,6] +
     NEW_DATA[,7]+ NEW_DATA[,8] + NEW_DATA[,9]+
     NEW_DATA[,10]+ NEW_DATA[,11]+ NEW_DATA[,12]+
     NEW_DATA[,13]+ NEW_DATA[,14]+ NEW_DATA[,15]+
                            ...
     NEW_DATA[,111]+ NEW_DATA[,112]+ NEW_DATA[,113]+
     NEW_DATA[,114]+ NEW_DATA[,115]+ NEW_DATA[,116]+
     NEW_DATA[,117]+ NEW_DATA[,118]+ NEW_DATA[,119]+
     NEW_DATA[,120]+ NEW_DATA[,121]+ NEW_DATA[,122]+
     NEW_DATA[,123]+ NEW_DATA[,124]+ NEW_DATA[,125])

i <-i+1
nn <-nn+1
ii <-ii+1
```

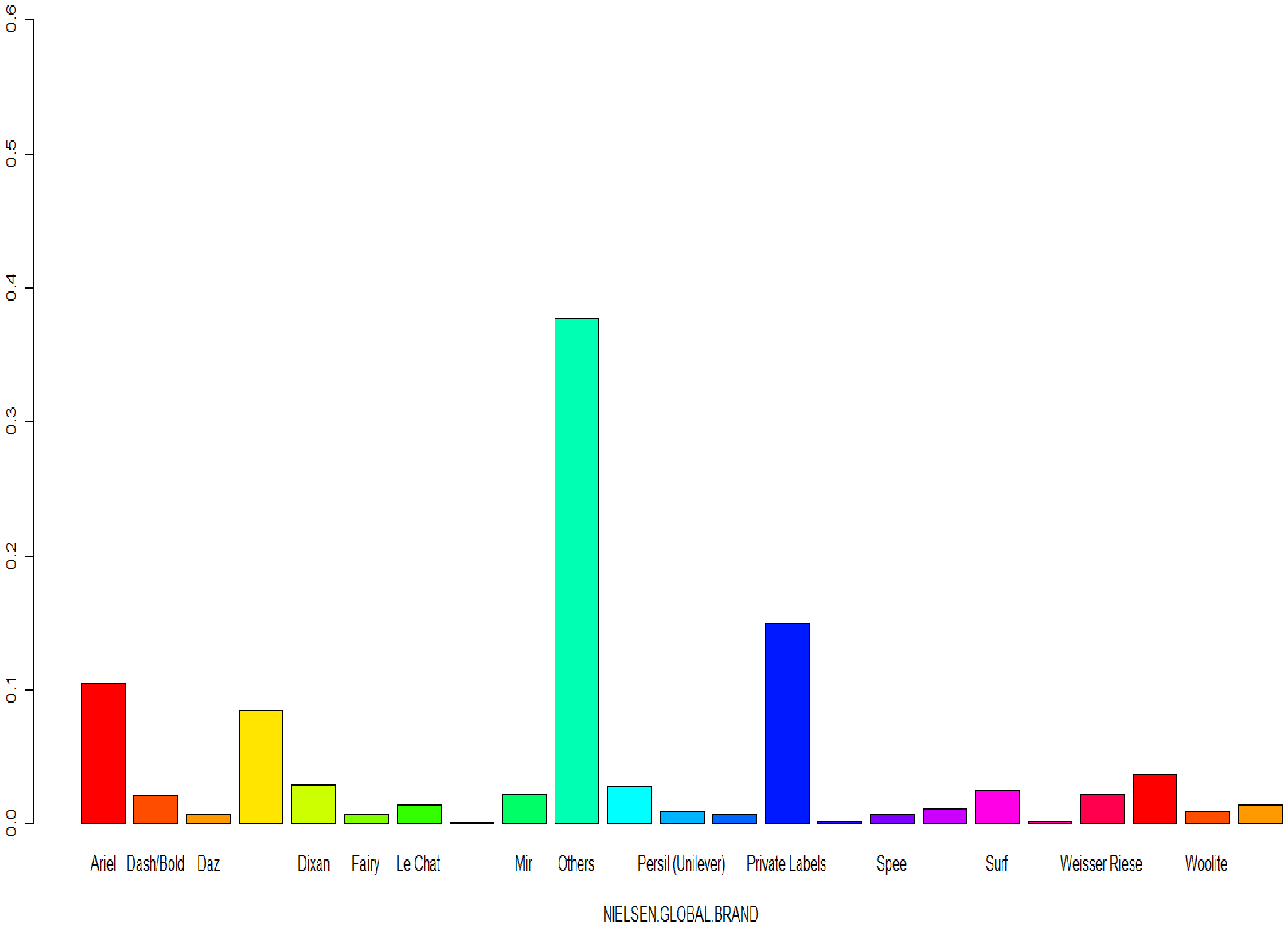
Here we start doing the predictions at each iteration for 705 products and storing those predictions in a list.

```
MSAM_predict <- predict(MSAM)

KAL[[nn]]<-MSAM_predict

   }
```

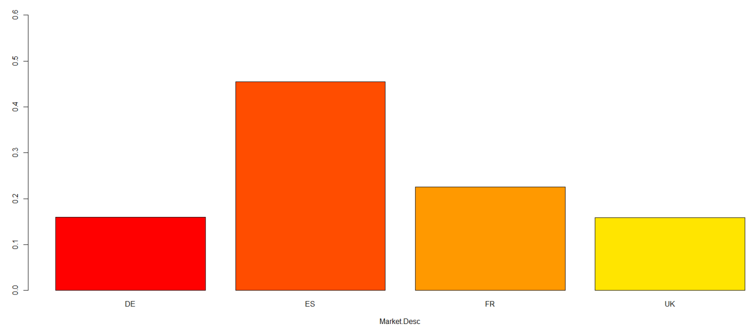
Here we start by unlisting the prediction values for 11 iterations and then calculating the MAPE and RMSPE value for those predicted values. In the end, those MAPE and RMSPE values are stored in two different lists.

```
YOYO <-as.data.frame(kal[1:705])
NONO <-as.data.frame(kal[706:1410])
POPO <-as.data.frame(kal[1411:2115])
QOQO <-as.data.frame(kal[2116:2820])
RORO <-as.data.frame(kal[2821:3525])
SOSO <-as.data.frame(kal[3526:4230])
TOTO <-as.data.frame(kal[4231:4935])
UOUO <-as.data.frame(kal[4936:5640])
VOVO <-as.data.frame(kal[5641:6345])
WOWO <-as.data.frame(kal[6346:7050])
MOMO <-as.data.frame(kal[7051:7755])

MSAM_COMBINED <-cbind(YOYO, NONO, POPO, QOQO, RORO, SOSO,
                      TOTO, UOUO, VOVO, WOWO, MOMO)

A1<-list()
A2<-list()

A1[[1]]<-MAE(MSAM_COMBINED$`kal[706:1410]`, NEW_DATA$Y_i29)
A2[[1]]<-RMSE(MSAM_COMBINED$`kal[706:1410]`, NEW_DATA$Y_i29)
A1[[2]]<-MAE(MSAM_COMBINED$`kal[1411:2115]`, NEW_DATA$Y_i30)
A2[[2]]<-RMSE(MSAM_COMBINED$`kal[1411:2115]`, NEW_DATA$Y_i30)
A1[[3]]<-MAE(MSAM_COMBINED$`kal[2116:2820]`, NEW_DATA$Y_i31)
A2[[3]]<-RMSE(MSAM_COMBINED$`kal[2116:2820]`, NEW_DATA$Y_i31)
A1[[4]]<-MAE(MSAM_COMBINED$`kal[2821:3525]`, NEW_DATA$Y_i32)
A2[[4]]<-RMSE(MSAM_COMBINED$`kal[2821:3525]`, NEW_DATA$Y_i32)
A1[[5]]<-MAE(MSAM_COMBINED$`kal[3526:4230]`, NEW_DATA$Y_i33)
A2[[5]]<-RMSE(MSAM_COMBINED$`kal[3526:4230]`, NEW_DATA$Y_i33)
```

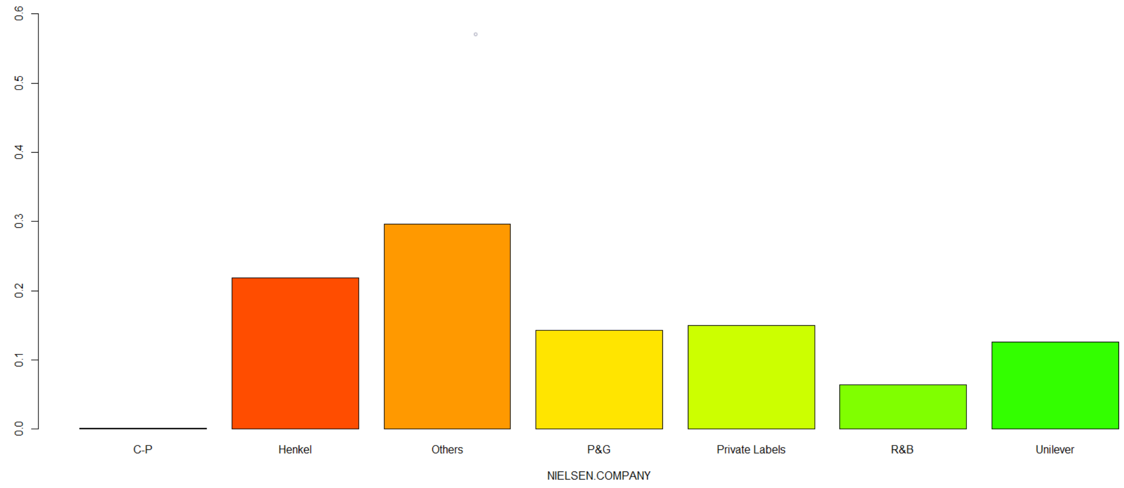
```r
A1[[6]]<- MAE(MSAM_COMBINED$`kal[4231:4935]`, NEW_DATA$Y_i34)
A2[[6]]<- RMSE(MSAM_COMBINED$`kal[4231:4935]`, NEW_DATA$Y_i34)
A1[[7]]<- MAE(MSAM_COMBINED$`kal[4936:5640]`, NEW_DATA$Y_i35)
A2[[7]]<- RMSE(MSAM_COMBINED$`kal[4936:5640]`, NEW_DATA$Y_i35)
A1[[8]]<- MAE(MSAM_COMBINED$`kal[5641:6345]`, NEW_DATA$Y_i36)
A2[[8]]<- RMSE(MSAM_COMBINED$`kal[5641:6345]`, NEW_DATA$Y_i36)
A1[[9]]<- MAE(MSAM_COMBINED$`kal[6346:7050]`, NEW_DATA$Y_i37)
A2[[9]]<- RMSE(MSAM_COMBINED$`kal[6346:7050]`, NEW_DATA$Y_i37)
A1[[10]]<-MAE(MSAM_COMBINED$`kal[7051:7755]`, NEW_DATA$Y_i38)
A2[[10]]<-RMSE(MSAM_COMBINED$`kal[7051:7755]`, NEW_DATA$Y_i38)

A1 <-unlist(A1)
A2 <-unlist(A2)

MAPE_RMSPE_MSAM <-cbind(A1, A2)
write.table(RMSPE_MSAM, "C:/Users/320078573/Desktop
             /PREDICT_MSAM", sep="\t")
```



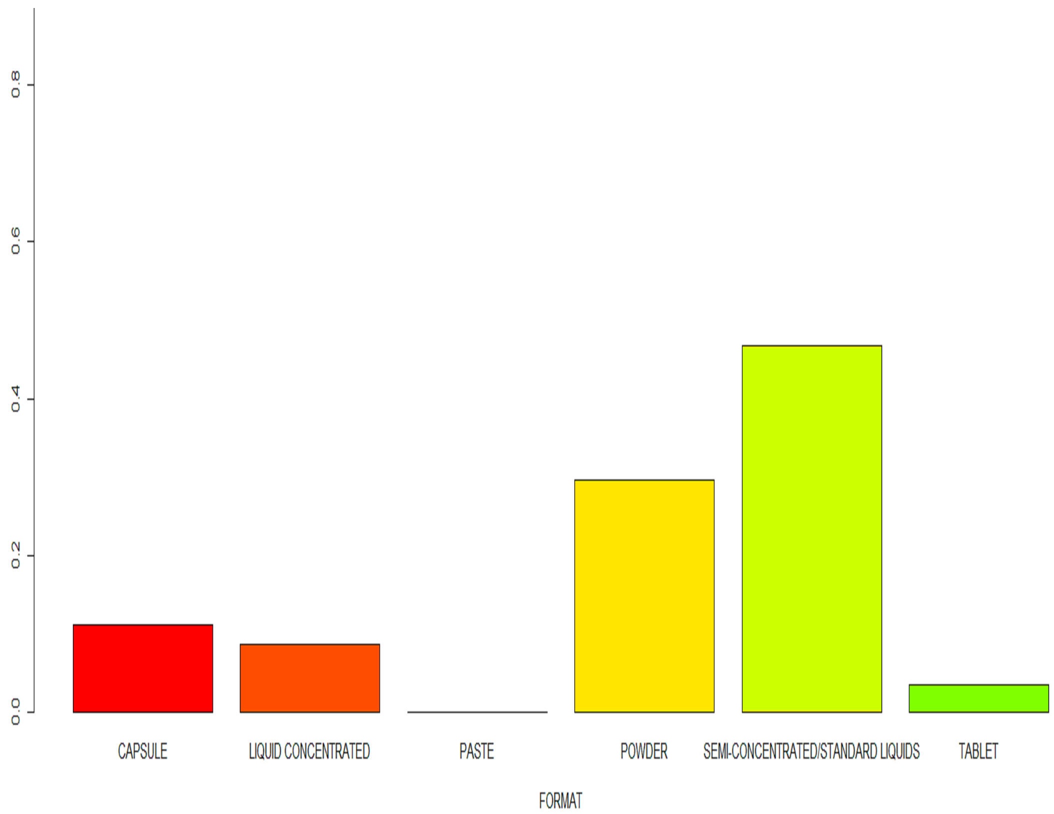
(c) Global Brand



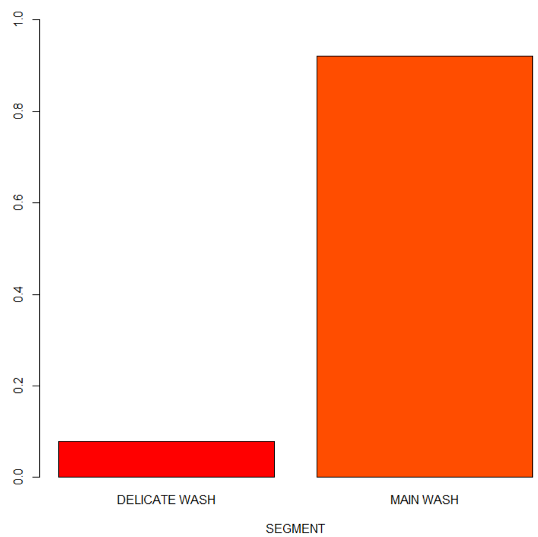
(d) Countries



(e) Company



(f) Format



(g) Segment

Figure 3: Statistics

Table A1: **Market share**

Variables	Min	Max	Mean	Median
Market Share 16P08	0.00000	44.60714	0.141844	0.00740
Market Share 16P09	0.00000	39.21009	0.141844	0.00304
Market Share 16P10	0.00000	38.63364	0.141844	0.00550
Market Share 16P11	0.00000	24.33068	0.141844	0.00381
Market Share 16P12	0.000001	19.54688	0.141844	0.00660
Market Share 16P13	0.00000	17.22942	0.141844	0.00533
Market Share 17P01	0.000001	23.55975	0.141844	0.00888
Market Share 17P02	0.000001	27.38004	0.141844	0.00967
Market Share 17P03	0.000001	18.81160	0.141844	0.00743
Market Share 17P04	0.00011	37.65036	0.141844	0.00588
Market Share 17P05	0.00018	14.54210	0.141844	0.00755
Market Share 17P06	0.00008	20.39587	0.141844	0.00284
Market Share 17P07	0.00009	23.95819	0.141844	0.00324
Market Share 17P08	0.00025	10.35240	0.141844	0.00929
Market Share 17P09	0.00012	35.80144	0.141844	0.00371
Market Share 17P10	0.00019	22.78281	0.141844	0.01204
Market Share 17P11	0.00033	15.51872	0.141844	0.00997
Market Share 17P12	0.00006	71.95930	0.141844	0.00319
Market Share 17P13	0.00008	71.68442	0.141844	0.00241
Market Share 18P01	0.00106	18.19042	0.141844	0.00324
Market Share 18P02	0.00040	11.92023	0.141844	0.01072
Market Share 18P03	0.00012	30.91826	0.141844	0.00558
Market Share 18P04	0.00002	79.09374	0.141844	0.00271
Market Share 18P05	0.00012	10.52687	0.141844	0.00881
Market Share 18P06	0.00013	32.90794	0.141844	0.00518
Market Share 18P07	0.00002	23.49893	0.141844	0.00838
Market Share 18P08	0.00002	22.89271	0.141844	0.00765
Market Share 18P09	0.00000	41.37816	0.141844	0.00305
Market Share 18P10	0.00000	53.04210	0.141844	0.00378
Market Share 18P11	0.00050	11.31902	0.141844	0.00954
Market Share 18P12	0.00000	68.45551	0.141844	0.00121
Market Share 18P13	0.000017	18.20106	0.141844	0.003717
Market Share 19P01	0.000013	17.51953	0.141844	0.004495
Market Share 19P02	0.00000	33.74230	0.141844	0.00217
Market Share 19P03	0.00000	40.72691	0.141844	0.00292
Market Share 19P04	0.000004	12.96296	0.141844	0.00754
Market Share 19P05	0.00000	70.48231	0.141844	0.00154
Market Share 19P06	0.00000	12.96640	0.141844	0.00422

Table A2: **Promo sales**

Variables	Min	Max	Mean	Median
Promo Sales 16P08	10	2088500	41683	210
Promo Sales 16P09	10	1379110	43266	340
Promo Sales 16P10	10	1700910	50373	500
Promo Sales 16P11	10	1366820	43999	450
Promo Sales 16P12	10	1074740	38531	340
Promo Sales 16P13	10	1332430	37620	600
Promo Sales 17P01	10	2065450	49469.5	1120
Promo Sales 17P02	10	2494010	96085	9990
Promo Sales 17P03	10	2972100	105531.1	12960
Promo Sales 17P04	10	2026960	86297	13094
Promo Sales 17P05	10	2637820	94338	13810
Promo Sales 17P06	10	2538910	89558	11570
Promo Sales 17P07	10	1986470	95857	12540
Promo Sales 17P08	10	1963020	78114	11710
Promo Sales 17P09	10	2753700	90870	10280
Promo Sales 17P10	10	991990	90668	10750
Promo Sales 17P11	10	1737830	81493	11140
Promo Sales 17P12	10	1901690	76595	10060
Promo Sales 17P13	10	1674910	78566	8570
Promo Sales 18P01	3.8	1836780	91117	11080
Promo Sales 18P02	10	2740570	83923	10594
Promo Sales 18P03	10	1950790	77721	9132
Promo Sales 18P04	10	1946880	77093	8500
Promo Sales 18P05	10	1810210	64644.8	7520
Promo Sales 18P06	10	2693530	62864	6600
Promo Sales 18P07	10	1845730	59125.2	6580
Promo Sales 18P08	11.3	2122830	50474.3	5807
Promo Sales 18P09	10	2233230	53485.8	5200
Promo Sales 18P10	10	3300250	58785	5460
Promo Sales 18P11	10	2063370	43946.4	4164.5
Promo Sales 18P12	10	1350840	40187	2960
Promo Sales 18P13	10	1723610	39183.8	2210
Promo Sales 19P01	10	2118520	48879	3360
Promo Sales 19P02	7.2	2606590	46391	2940
Promo Sales 19P03	10	2089980	46281	2331
Promo Sales 19P04	10	1451210	35209.4	1900
Promo Sales 19P05	7.3	1854680	39901.5	2040
Promo Sales 19P06	7	3269890	45605	1580

Table A3: **Volume sales**

Variables	Min	Max	Mean	Median
Volume Sales 16P08	50	102044710	1628460	439020
Volume Sales 16P09	70	95070030	1625729	433030
Volume Sales 16P10	90	100475230	1654237	448790
Volume Sales 16P11	130	96307660	1577167	443720
Volume Sales 16P12	560	95118930	1525826	429650
Volume Sales 16P13	720	99375580	1509634	424760
Volume Sales 17P01	770	92710520	1602737	408960
Volume Sales 17P02	550	99140480	1536640	408110
Volume Sales 17P03	60	101806520	1652363	412670
Volume Sales 17P04	740	107792780	1571214	446010
Volume Sales 17P05	930	113344020	1676398	434640
Volume Sales 17P06	1100	121218710	1665413	468420
Volume Sales 17P07	1260	115067460	1679645	467760
Volume Sales 17P08	1300	109841370	1604687	458610
Volume Sales 17P09	1310	108014260	1637261	461880
Volume Sales 17P10	970	102044710	1642271	415630
Volume Sales 17P11	1020	95070030	1525041	418440
Volume Sales 17P12	1020	100475230	1507952	393100
Volume Sales 17P13	850	96307660	1450898	400970
Volume Sales 18P01	700	95118930	1526847	402730
Volume Sales 18P02	720	99375580	1458082	362850
Volume Sales 18P03	770	92710520	1451857	354500
Volume Sales 18P04	840	90971200	1475743	361560
Volume Sales 18P05	830	90282470	1383533	313450
Volume Sales 18P06	640	94000040	1388148	300060
Volume Sales 18P07	890	94451070	1373553	279830
Volume Sales 18P08	960	94207430	1293885	262330
Volume Sales 18P09	730	89090170	1244017	234200
Volume Sales 18P10	160	85896690	1237677	214770
Volume Sales 18P11	620	77692330	1142773	201160
Volume Sales 18P12	630	80792550	1133616	184770
Volume Sales 18P13	570	79475220	1121406	166050
Volume Sales 19P01	590	81580010	1151747	163050
Volume Sales 19P02	80	87086490	1097376	153400
Volume Sales 19P03	150	87331230	1128425	144150
Volume Sales 19P04	90	84374500	1061284	141060
Volume Sales 19P05	140	92269980	1090752	136830
Volume Sales 19P06	40	95669880	1076220	130150